Evolution of phage-type RNA polymerases in higher plants

DISSERTATION

Zur Erlangung des akademischen Grades
Doctor rerum naturalium (Dr. rer. nat.) im Fach Biologie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
der Humboldt-Universität zu Berlin

von
Master der Landwirtschaftswissenschaften
Chang Yin

Präsident der Humboldt-Universität zu Berlin
Prof. Dr. h.c. Christoph Markschies

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I
Prof. Dr. Andreas Herrmann

Gutachter: 1. Prof. Dr. Thomas Börner

2. Prof. Dr. Volker Knoop

3. Prof. Dr. Wolfgang Schuster

Datum der Promotion: 29.11.2010

Table of contents

Zusammenfassung

In mono- und eudikotylen Pflanzen kodiert eine Genfamilie (*RpoT*, RNA-Polymerase des *T3/T7*-Typs) mitochondriale und plastidäre RNA-Polymerasen (RNAP), die den ungeraden T-Phagen-Polymerasen ähneln. *RpoT*-Gene von Angiospermen sind gut charakterisiert, während aus tiefer abzweigenden Pflanzenspecies bisher lediglich die Gene aus dem Moos *Physcomitrella* beschrieben wurden. Um einen Beitrag zur Aufklärung der molekularen Evolution der RpoT-Polymerasen im Pflanzenreich zu liefern und um Erkenntnisse über die potentielle Bedeutung von multiplen Phagen-Typ (RNAP) in Pflanzen zu gewinnen, wurden die *RpoT*-Gene aus dem Lycophyten *Selaginella moellendorffii* und aus dem basalen Angiosperm *Nuphar advena* identifiziert und charakterisiert.
Selaginella moellendorffii (Moosfarn)-*Trace*-Sequenzdaten mit hoher Ähnlichkeit zu *RpoT*-Sequenzen von Angiospermen wurden benutzt, um das *full-length SmRpoT*-Gen und die entsprechende cDNA zu isolieren. Die *SmRpoT*-mRNA ist 3542 nt lang und weist einen offenen Leserahmen von 3006 nt auf, der für ein putatives Protein aus 1002 Aminosäuren mit einer molekularen Masse von 113 kDa kodiert. Das *SmRpoT*-Gen besteht aus 19 Exons und 18 Introns, die in ihren Positionen mit denen aus den Angiosperm- und *Physcomitrella*-Genen konserviert sind. Mittels Southernblot-Analyse wurde nachgewiesen, dass *S. moellendorffii* ein *single-copy RpoT*-Gen kodiert. Für das N-terminale Transitpeptid von SmRpoT konnte gezeigt werden, dass es bei transienter Expression in *Arabidopsis*- und *Selaginella*-Protoplasten den Transport von GFP (*green fluorescent protein*) exclusiv in Mitochondrien vermittelt. In *N. advena* wurden mittels Screening einer BAC-Bibliothek drei *RpoT*-Gene identifiziert. Sowohl die genomischen als auch die cDNA-Sequenzen wurden aufgeklärt. Die *NaRpoT*-mRNAs kodieren putative Polypeptide von 996, 990 und 985 Aminosären. Alle drei Gene besitzen 19 Exons und 18 Introns, die in ihren Positionen mit denen der *RpoT*-Gene aus *Selaginella* und allen anderen Landpflanzen konserviert sind. Die kodierten Proteine weisen auf Aminosäureebene einen hohen Konservierungsgrad auf, einschließlich aller essentiellen Regionen und Aminosäurereste, die für die T7-RNAP bekannt sind. Die N-terminalen Transitpeptide zweier der kodierten RNAP, NaRpoTm1 und NaRpoTm2, vermittelten den Import von GFP exclusiv in Mitochondrien, während die dritte Polymerase, NaRpoTp, in Chloroplasten importiert wurde. Interessanterweise muß die Translation der *NaRpoTp*-mRNA an einem CUG-Codon initiiert werden, um ein funktionelles Protein mit plastidärem Transitpeptid zu erhalten. Die *N. advena* RpoTp-RNAP ist somit neben *AGAMOUS* aus *Arabidopsis* und der RpoTp-RNAP aus *Nicotiana*, ein weiteres Beispiel für jene selten vorkommenden pflanzlichen mRNAs, deren Translation exclusiv an nicht-AUG-Codons initiiert wird.
Die Rekonstruktion von phylogenetischen Bäumen resultierte in unterschiedlichen Positionen für die *Selaginella*- und *Nuphar*-Polymerasen: Im Gegensatz zu der RpoT-Polymerase aus *S. moellendorffii* und denen aus *Physcomitrella*, die in den phylogenetischen Analysen Schwesterpositionen zu allen anderen Phagentyp-RNAP der Angiospermen einnehmen, clusterten die *Nuphar*-RpoTs zusammen mit den deutlich separierten mitochondrialen (NaRpoTm1 und NaRpoTm2) und plastidären (NaRpoTp) Polymerasen. *Selaginella* kodiert eine einzige mitochondriale RNAP, während *Nuphar* zwei mitochondriale und eine plastidäre RNAP besitzt. Die Identifizierung einer Plastiden-lokalisierten Phagentyp-RNAP in diesem basalen Eudikotylen, die ortholog zu allen anderen RpoT-Enzymen der Blütenpflanzen ist, läßt darauf schließen, daß die Acquisition einer nukleär kodierten plastidären RNAP, die noch in den Lycopoden fehlt, nach der Trennung der Leucopoden von allen anderen Tracheophyten erfolgte. Eine “dual-targeting” RNAP (mitochondrial *und* plastidär lokalisiert), wie sie in Eudikotylen, nicht jedoch in Monokotylen vorkommt, wurde weder in *Selaginella* noch in *Nuphar* nachgewiesen, vermutlich ist sie ein evolutionäres Novum von eudikotylen Pflanzen wie *Arabidopsis*.

Summary

In mono- and eudicot plants, a small nuclear gene family (*RpoT, R*NA *po*lymerase of the *T*3/*T*7 type) encodes mitochondrial as well as chloroplast RNA polymerases homologous to the T-odd bacteriophage enzymes. *RpoT* genes from angiosperms are well characterized, whereas data from deeper branching plant species until recently were limited to the moss *Physcomitrella.* To elucidate the molecular evolution of the RpoT polymerases in the plant kingdom and to get more insight into the potential importance of having more than one phage-type RNA polymerase (RNAP) available, we identified and characterized *RpoT* genes in the lycophyte *Selaginella moellendorffii* and the basal eudicot *Nuphar advena*.

Selaginella moellendorffii (spikemoss) sequence trace data encoding a polypeptide highly similar to angiosperm and moss phage-type organelle RNA polymerases were used to isolate a BAC clone containing the full-length gene *SmRpoT* as well as the corresponding cDNA. The *SmRpoT* mRNA comprises 3452 nt with an open reading frame of 3,006 nt, encoding a putative protein of 1,002 amino acids with a molecular mass of 113 kDa. The *SmRpoT* gene comprises 19 exons and 18 introns, conserved in their position with those of the angiosperm and *Physcomitrella RpoT* genes. Using Southern blot analysis, it was shown that *S. moellendorffii* encodes a single *RpoT* gene. The N-terminal transit peptide of SmRpoT was shown to confer targeting of green fluorescent protein (GFP) exclusively to mitochondria after transient expression in *Arabidopsis* and *Selaginella* protoplasts.

In *Nuphar advena* three *RpoT* genes were identified by BAC library screening. Both genomic gene sequences and full-length cDNAs were determined. The *NaRpoT* mRNAs specify putative polypeptides of 996, 990 and 985 amino acids, respectively. All three genes comprise 19 exons and 18 introns, conserved in their positions with those from *S. moellendorffii* and the *RpoT* genes of other land plants. The encoded proteins show a high degree of conservation at the amino acid sequence level, including all functional crucial regions and residues known from the phage T7 RNAP. The N-terminal transit peptides of two of the encoded polymerases, NaRpoTm1 and NaRpoTm2, conferred targeting of GFP exclusively to mitochondria, whereas the third polymerase, NaRpoTp, was targeted to chloroplasts. Remarkably, translation of *NaRpoTp* mRNA has to be initiated at a CUG codon to generate a functional plastid transit peptide. Thus, besides AGAMOUS in *Arabidopsis* and the *Nicotiana* RpoTp polymerase, *N. advena RpoTp* provides another example for a plant mRNA that is exclusively translated from a non-AUG codon.

Reconstruction of phylogenetic trees revealed different positions of the RpoTs from the lycophyte *Selaginella* and the basal eudicot *Nuphar*. In contrast to the RpoTs of *S. moellendorffii* and those of the moss *Physcomitrella*, which are according to the phylogenetic analyses in sister positions to all other phage-type polymerases of angiosperms, the *Nuphar* RpoTs clustered with the well separated clades of mitochondrial (NaRpoTm1 and NaRpoTm2) and plastid (NaRpoTp) polymerases.

Selaginella encodes a single mitochondrial RNAP, whereas *Nuphar* harbors two mitochondrial and one plastid phage-type polymerases. Identification of a plastid localized phage-type RNAP in this basal eudicot, orthologous to all other RpoTp enzymes of flowering plants, suggests that the acquisition of a nuclear encoded plastid RNA polymerase, not present in lycopods, took place after the split of lycopods from all other tracheophytes. A dual-targeted mitochondrial and plastid RNA polymerase (RpoTmp), as present in eudicots but not monocots, was not detected in *Nuphar* or *Selaginella* suggesting that its occurrence is an evolutionary novelty of eudicotyledoneous plants like *Arabidopsis*.

Schlagwörter: *Selaginella moellendorffii*, *Nuphar advena*, Phage-Typ RNA Polymerasen, Gene Duplikation

Keywords: *Selaginella moellendorffii*, *Nuphar advena*, phage-type RNA polymerase, gene duplication

1 Introduction

1.1 Organellar RNA polymerases in plants and their functions

Plastids and mitochondria were acquired through endosymbiosis from ancestral free-living cyanobacteria and α-proteobacteria, respectively (Gray, 1989; Gray, 1992; Gray, 1993). They possess their own genomes and transcriptional systems. Recent work on organellar transcription in plants has shed new light onto the molecular transcriptional machineries of plastids and mitochondria (Hess & Borner, 1999; Weihe, 2004).

1.1.1 Mitochondrial RNA polymerases

The core-subunit of the mitochondrial RNA polymerase (RNAP) in most eukaryotic organisms is represented by a single-subunit RNAP that is homologous to the RNA polymerases of the bacteriophages T3 and T7 that are, encoded by nuclear *RpoT* (T = T phage-like) genes (Hess & Borner, 1999). For the protozoan *Reclinomonas americana*, the mitochondrial transcription relies on a chondriom-encoded multi-subunit proteobacterial RNAP (Lang *et al.*, 1997). However, all eukaryotes seem to have the nuclear encoded phage-type transcription machinery.

Mitochondrial-targeted RpoT polymerases have been identified in numerous dicotyledonous and monocotyledonous plants (Weihe *et al.*, 1997; Chang *et al.*, 1999; Ikeda & Gray, 1999; Kobayashi *et al.*, 2001b). Evidence for a nucleus-encoded RNA polymerase in plant mitochondria is more indirect. One of the first findings was that in the mitochondrial genomes sequenced to date, no open reading frames were found encoding homologs of known RNAP subunits (Oda *et al.*, 1992; Unseld *et al.*, 1997), meaning the mitochondrial RNAP of plants and algae must be encoded in the nucleus. A second finding was that plant mitochondrial promoters are shown to resemble those of yeast mitochondria (Rapp & Stern, 1992; Binder *et al.*, 1995; Caoile & Stern, 1997). A third finding-, was that genes from nucleus-encoded T7-like RNA polymerases are widely involved in mitochondrial transcription in eukaryote systems (Cermakian *et al.*, 1996), including human (Tiranti *et al.*, 1997), yeast (*Saccharomyces cerevisiae*) (Greenleaf *et al.*, 1986; Kelly *et al.*, 1986; Masters *et al.*, 1987), *Neurospora crassa* (Chen *et al.*, 1996) and *Schizosaccharomyces pombe* (accession No. Z99126), as well as in plants, such as *Arabidopsis* (Hedtke *et al.*, 1997), *Chenopodium* (Weihe *et al.*, 1997), wheat (Ikeda & Gray, 1999), rice (Cermakian *et al.*, 1996) and maize (Young *et al.*, 1998).

1.1.2 Plastid RNA polymerases

Plastids harbor two different types of RNAP: a eubacterial-type multi-subunit enzyme (PEP = **p**lastid-***e***ncoded plastid RNA **p**olymerase), and a single-subunit core RNAP (NEP = **n**uclear-**e**ncoded RNA **p**olymerase). In land plants and algae, most plastid genes are transcribed by a eubacterial-type plastid-encoded RNA polymerase (PEP) that is encoded by plastids and can be traced back to the plastid endosymbiont (Igloi & Kossel, 1992; Hess & Borner, 1999). PEP are composed of four subunits (α, β, β', β'', shown in Tab. 1) and one of many nuclear-encoded accessory transcription factors, such as sigma-like factors (Tanaka *et al.*, 1996; Tanaka *et al.*, 1997; Isono *et al.*, 1997; Oikawa *et al.*, 1998; Kestermann *et al.*, 1998) or DNA binding proteins(Kim & Mullet, 1995; Trifa *et al.*, 1998). The subunit structure of the core RNAP complex is homologous to the multi-subunit RNAPs of cyanobacteria, which is an indication of the endosymbiotic origin of plastids from cyanobacteria-like prokaryotes.

Tab. 1 Eubacterial-type plastid-encoded RNA polymerase (PEP)

Gene		Encoding RNAP subunit	Mass weight
rpoA		A	38 KDa
rpoB		B	120 KDa
rpoC	rpoC1	β'	85 KDa
	rpoC2	β''	185 KDa

Additionally, plastids of eudicots, monocots and the moss *Physcomitralla* harbor a phage-type RNA polymerase (NEP) that is encoded by nuclear *RpoT* gene(s) in a strategy similar to the mitochondrial RNAP. Evidence for the existence of a nucleus-encoded NEP came from studies where chloroplast translation (and synthesis of PEP as well) was blocked or PEP was directly inactivated (Kapoor *et al.*, 1997) but this did not lead to a complete inhibition of transcription, suggesting that residual transcription resulted from RpoT polymerases (Hess *et al.*, 1993; Silhavy & Maliga, 1998). It was revealed that NEP activity resulting from a T7-like RNA polymerase was present in spinach chloroplasts (Lerbsmache, 1993). Indeed, the transit peptide of a T7-type RNA polymerase from *Arabidopsis*, when fused to the green fluorescent protein (GFP), can directly import the GFP into spinach chloroplasts *in vitro* (Hedtke *et al.*, 1997) and to *Arabidopsis* chloroplasts *in vivo* (Hedtke *et al.*, 1999). Therefore it was concluded that a nuclear-encoded T7-like RNA polymerase contributes to chloroplast transcription in higher plants.

1.1.3 *RpoT* genes encode both mitochondrial and plastid RNAPs in angiosperms

Nuclear genomes of higher plants harbor a small family of genes encoding for organellar RNA polymerases, that are homologous to and related to the RNAPs from the bacteriophages T3, T7, K11 and SP6 (Weihe, 2004), and are widely distributed among the eukaryote lineage (Cermakian *et al.*, 1997).

In accordance the Commission on Plant Gene Nomenclature, and in order to easily discriminate them from chloroplast genes encoding RNAP subunits (*rpoA*, *B*, *C1*, *C2*), the designation '*RpoT*' genes (RNA polymerases of the T3/T7 phage single subunit type) were suggested for the plant genes encoding phage-type RNAPs, where 'T' indicates the relationship to the T3- and T7-type RNAPs.

A number of nuclear *RpoT* genes have been identified in different species, including humans (Tiranti *et al.*, 1997), fungi (Masters *et al.*, 1987; Chen *et al.*, 1996), *Plasmodium falciparum* (Li *et al.*, 2001), *Chenopodium album* (Weihe *et al.*, 1997), *Arabidopsis thaliana* (Hedtke *et al.*, 1997; Hedtke *et al.*, 1999; Hedtke *et al.*, 2000), maize (Young *et al.*, 1998; Chang *et al.*, 1999), wheat (Ikeda & Gray, 1999), *Nicotiana tabacum* (Hedtke *et al.*, 2002), *Nicotiana sylvestris* (Kobayashi *et al.*, 2001a; Kobayashi *et al.*, 2002; Hedtke *et al.*, 2002), barley (Emanuel *et al.*, 2004) and rice (Kusumi *et al.*, 2004). The *RpoT* genes from plants are briefly summarized in Tab. 2.

Tab. 2 Overview of nuclear RpoT genes in plants

Species	*RpoT* gene	Localization of encoded protein	Reference
Chenopodium album	CaRpoT	mitochondria	Weihe et al., 1997; Hedtke et al., 1999.
Physcomitrella patens	PpRpoT1	dually targeting or mitochondria	Kabeya et al., 2002; Richter et al., 2002
	PpRpoT2		Richter et al., 2002
	PpRpoT3	not investigated	the Physcomitrella genome project
Arabidopsis thaliana	RpoT1 (RpoTm)	mitochondria	Cermakian et al., 1996; Hedtke et al., 1997; Emanuel et al., 2006
	RpoT2 (RpoTmp)	dually targeting	Hedtke et al., 2000; Emanuel et al., 2006
	RpoT3 (RpoTp)	plastid	Chang et al., 1999; Ikeda and Gray, 1999; Hedtke et al., 1997, 1999; Emanuel et al., 2006
Nicotiana sylvestris	NsRpoT-A	mitochondria	Kobayashi et al., 2001a.

	NsRpoT-B	dually targeting	Kobayashi et al., 2001b.
	NsRpoT-C	plastid	Kobayashi et al., 2002.
Nicotiana tabacum	2x RpoT1	mitochondria	Hedtke et al., 2002
	2x RpoT2	dually targeting	
	2x RpoT3	plastid	
Barley	RpoTm	mitochondria	Emanuel et al., 2004
	RpoTp	plastid	
Wheat	wheat-G/ RpoTm	mitochondria	Ikeda and Gray, 1999
	wheat-C	plastid	
Maize	rpoTm	mitochondria	Young et al., 1998; Chang et al., 1999
	rpoTp	plastid	
Rice	OsRpoTp	plastid	Kusumi et al., 2004
	OsRpoTm	mitochondria	
Cucumis sativus	3xRpoT	not investigated	Huang et al., 2009
Populus trichocarpa	PotRpoT1	not investigated	data from genome project
	PotRpoT2		
	PotRpoT3		

Analysis of the sub-cellular localization of *RpoT* gene products led to the identification of the plant phage-type RNA polymerases. *RpoTm* targeting the mitochondria, and *RpoTp* targeting the plastids (Fig. 1). In the majority of monocotyledonous plants, two *RpoT* genes were indentified (Hedtke *et al.*, 1997; Hedtke *et al.*, 2002; Chang *et al.*, 1999; Ikeda & Gray, 1999; Emanuel *et al.*, 2004). An exception to this was the moss *Physcomitrella patens* that has three *RpoT* genes (Kabeya *et al.*, 2002). Conversely, all dicotyledoneous plants analyzed to date, have three *RpoT* genes. The polymerase encoded by the third gene is hypothesized to encode a protein that likely targets both mitochondria and plastids and is designated *RpoTmp* (Hedtke *et al.*, 2000; Hedtke *et al.*, 2002; Kobayashi *et al.*, 2001a). Therefore, in *Arabidopsis*, both mitochondria and plastids have two types of RpoT enzymes (i.e., RpoTm and RpoTmp for mitochondria, RpoTp and RpoTmp for chloroplast). Subcellular localization was demonstrated in many cases by GFP fusion constructs and/or *in vitro* import studies (Hedtke *et al.*, 1999; Richter *et al.*, 2002). With the bacterial-type PEP and two different

nuclear-encoded phage-type RNAP (RpoTp & and RpoTmp), *Arabidopsis* (and other dicots) plastids possess three different organellar transcriptional activities.

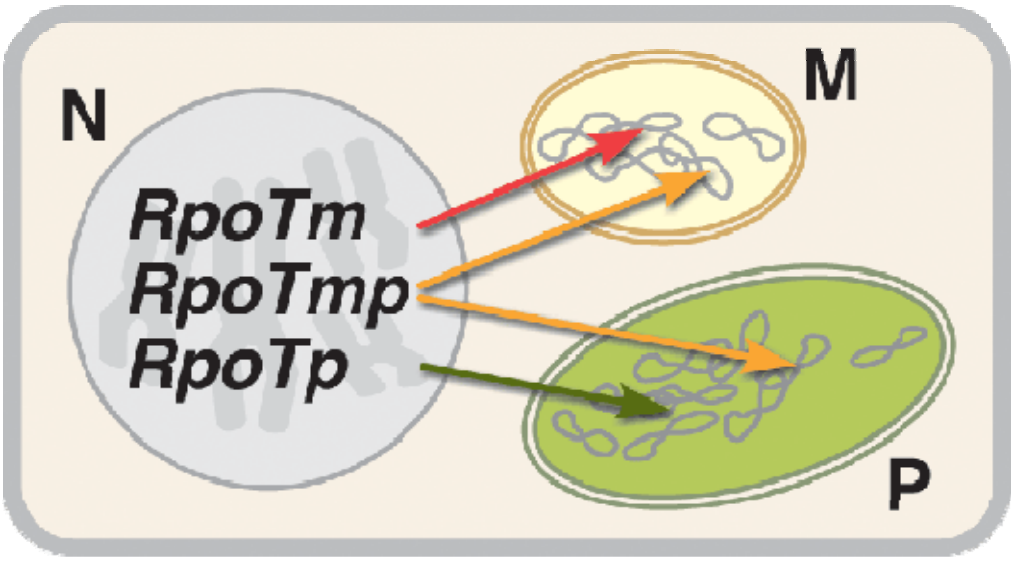


Fig. 1 Three phage-type RNA polymerases (RpoTm, RpoTp and RpoTmp) are encoded by the nuclear (N) genome of eudicots (http://www2.hu-berlin.de/ biologie/sfb429/tp_a5.html). They have been confirmed by bioinformatic and experimental approaches to be targeted to mitochondria (M), to plastids (P), and to both organelles (M and P), respectively.

In the moss *Physcomitrella patens*, three *RpoT* genes were identified (Kabeya *et al.*, 2002; Richter *et al.*, 2002). Two of them were shown to be capable of dual-targeting to mitochondria and plastids, with multiple translation initiation as the underlying mechanism (Richter *et al.*, 2002).

All higher plant nuclear *RpoT* genes identified to date consist of 19 exons and 18 introns at conserved positions (Fig. 2). The *RpoT* genes of all angiosperms studied to date share high sequence homology and remarkable conservation of intron number and position. They can all be traced back to a common ancestor, indicating that the multiple genes resulted from gene duplication events and represent paralogs.

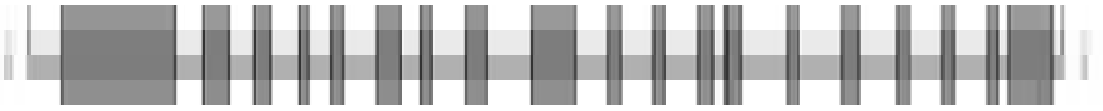


Fig. 2 Gene structure of RpoT (Modified from Emanuel et al., 2004). Translated and untranslated regions are indicated by filled boxes and bold lines, respectively.

1.1.4 Structure, promoter recognition and functions of RpoT in plants

1.1.4.1 Structure of RpoT

The catalytically active C-terminus of T7 RNA polymerase can be divided into 'fingers', 'palm' and 'thumb' domains (Cheetham *et al.*, 1999) as shown in fig. 3. However, the N-terminal regions diverge strongly, bearing elements responsible for promoter recognition. In yeast (*Saccharomyces cerevisiae*), the mitochondrial RNA polymerase (mtRNAP) is encoded by a nuclear gene (*Rpo41)* and is homologous to the single-polypeptide RNAPs of the bacteriophages T3 and T7 (Greenleaf *et al.*, 1986; Kelly *et al.*, 1986; Masters *et al.*, 1987). Two regions of yeast polymerase Rpo41, i.e., the 'specificity loop' and intercalating β-hairpin that are involved in promoter recognition and DNA melting, are conserved more in structure than sequence (Matsunaga & Jaehning, 2004a; Matsunaga & Jaehning, 2004b). It was suggested that the corresponding structures might also exist for plant RpoT polymerases, giving RpoT enzymes the intrinsic ability for preferential initiation of transcription at particular sequences *in vitro* (Kuhn *et al.*, 2007). The sequence similarity of RpoTs between plants, yeast and human increases from the NH_2 terminus to the COOH terminus, which is consistent with the catalytic function of the carboxy terminus of RNAPs.

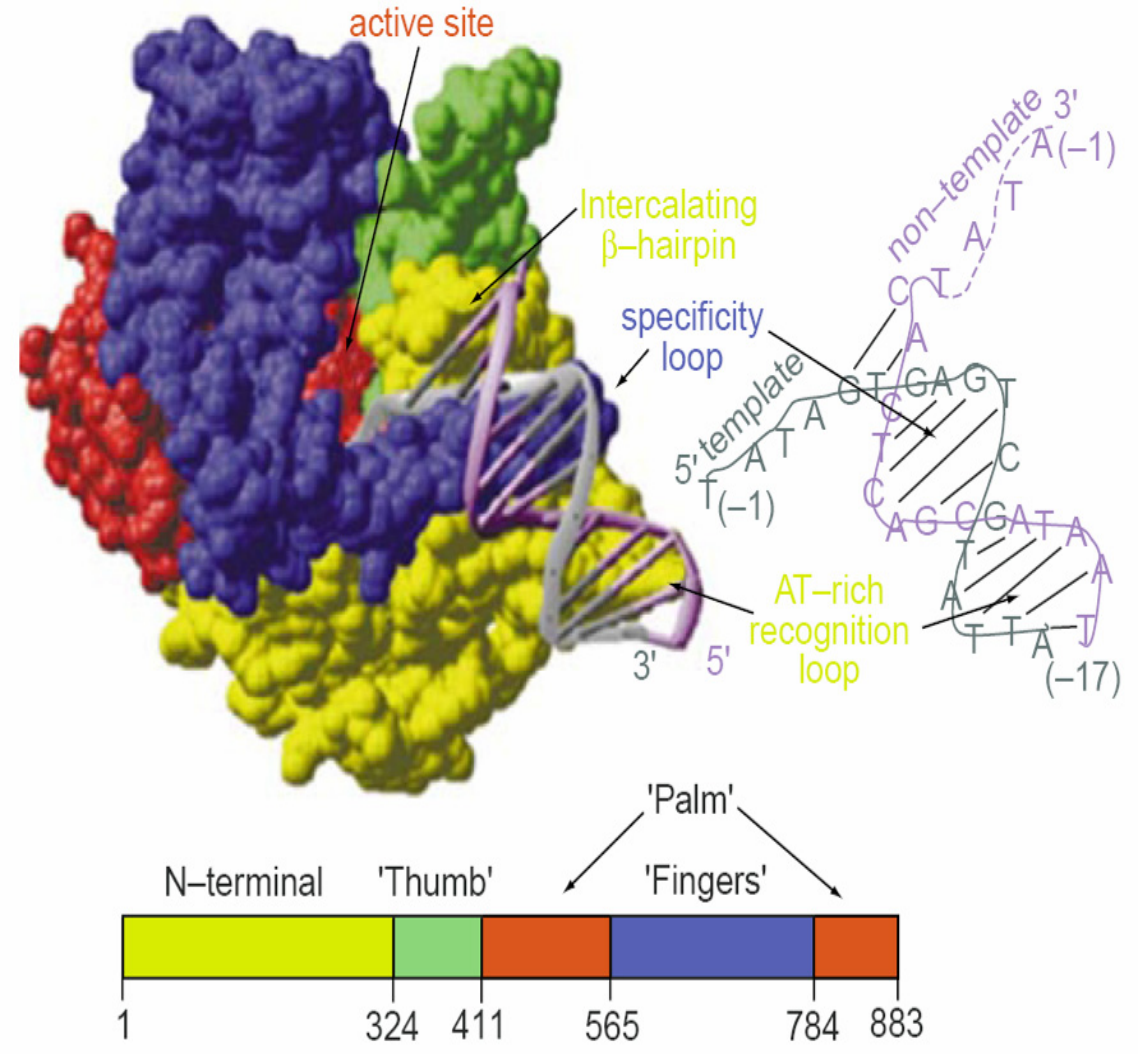


Fig. 3 Structure of the T7 RNAP-promoter complex (Cheetham et al., 1999). According to the sub-domains, the catalytically active C-terminus of T7 RNA polymerase can be divided into three parts: `palm' (red), `thumb' (green) and `fingers' (blue). The N-terminal domain (yellow) is specific to all phage-like RNA polymerases. The template (grey) and non-template (magenta) strands are shown as well. The sequence of the promoter and sequence organization of the T7 RNAP domains are also shown.

1.1.4.2 Promoter recognition by RpoT

Plant mitochondria possess numerous promoters for recognition by RpoTm (and RpoTmp). Frequently, multiple promoters are active in the upstream region of one gene (Lupold *et al.*, 1999; Kuhn *et al.*, 2005). Promoter sequences of up to 25 nt, comprising the transcription start site, are required for initiation of transcription *in vitro* (Rapp & Stern, 1992; Caoile & Stern, 1997; Dombrowski *et al.*, 1999). Many, but not all plant mitochondrial promoters contain a YRTA core (Weihe, 2004). There is limited data on the recognition of these promoters by different RNAPs. Similar to T7 polymerase and the yeast mitochondrial RNA, *Arabidopsis* RpoTm (and RpoTp) were shown to initiate transcription from several promoters *in vitro* and

catalyze transcript elongation without any additional co-factors, when promoters were added as super-coiled DNA templates (Kuhn *et al.*, 2007). In contrast to RpoTm, RpoTmp was active on very few sequences and at a very low efficiency, suggesting that RpoTm is the main mitochondrial RNAP. *In vivo*, the dual-targeted RpoTmp of Arabidopsis appears to describe a specific subset of mitochondrial genes. It has been shown that in RpoTmp mutants plants the levels of specific transcripts were correlated with reduced abundances of the respiratory chain complexes I and IV (Kuhn *et al.*, 2009). The decreased transcription, however, was not associated with changes in promoter utilization, suggesting that RpoTmp function is gene specific rather than promoter specific.

Transcription of plastid genes of angiosperms needs an additional nuclear-encoded RNA polymerase (NEP) in addition to the plastid-encoded PEP (Hess & Borner, 1999; Shiina *et al.*, 2005; Liere K & Borner T., 2007). As shown in Tab. 3, most NEP promoters consist of a core sequence (YRTA; type-Ia), similar to a number of plant mitochondrial promoters (Liere & Maliga, 1999; Weihe & Borner, 1999; Binder & Brennicke, 2003; Kuhn *et al.*, 2005). A subclass of NEP promoters shares a GAAbox motif upstream of the YRTA-motif (type-Ib), (Kapoor & Sugiura, 1999). Type-II NEP promoters, represented by dicot *clpP* promoters, lack these motifs and possess crucial sequences located downstream of the transcription initiation site (Weihe & Borner, 1999; Liere & Maliga, 1999).

Tab. 3 Plastid NEP promoters

NEP promoters	Motif	Characteristics	Reference
class Ia	core YRTA motif	Resembling mitochondrial promoters	Weihe and Börner, 1999
class Ib	additional GAA-box motif	upstream of the YRTA-motif, important in transcription from, e.g., the tobacco PatpB-289 NEP promoter	Kapoor and Sugiura, 1999; Weihe and Börner, 1999; Liere and Börner, 2007
Class II	Lack of the YRTA motif	crucial sequences located downstream of the transcription initiation site	Sriraman et al., 1998; Weihe and Börner, 1999

The existence of additional non-consensus NEP promoters (Pc) has been reported for the *rrn* operon in spinach, mustard, *Arabidopsis*, and for the internal promoters of certain tRNAs (Liere and Börner 2007). Activity inhibition of three *Arabidopsis* RpoT enzymes by three plastid tRNAs ($tRNA^{Glu}$, $tRNA^{Gly}$ and $tRNA^{val}$) was also observed (Bohne A-V *et al.*, 2009). To discriminate between NEP and PEP promoters, Swiatecka-Hagenbruch et al.

(Swiatecka-Hagenbruch *et al.*, 2007) compared the 5'-ends of transcripts from chlorophyll-deficient *Arabidopsis* plants with those of untreated plants. Using 5'-RACE combined with enzymatic treatment of RNAs to recognize primary and secondary 5'-ends, they unambiguously identified transcription initiation sites of the *Arabidopsis accD, atpB, atpI, rpoB, rps4, rps15*, and *ycf1* genes (Fig. 4). The diversity in individual promoter usage between plants suggests species-specific solutions for attaining control over gene expression in plastids, and although many genes are transcribed from a single promoter, transcription of plastidial genes and operons by multiple promoters appears common.

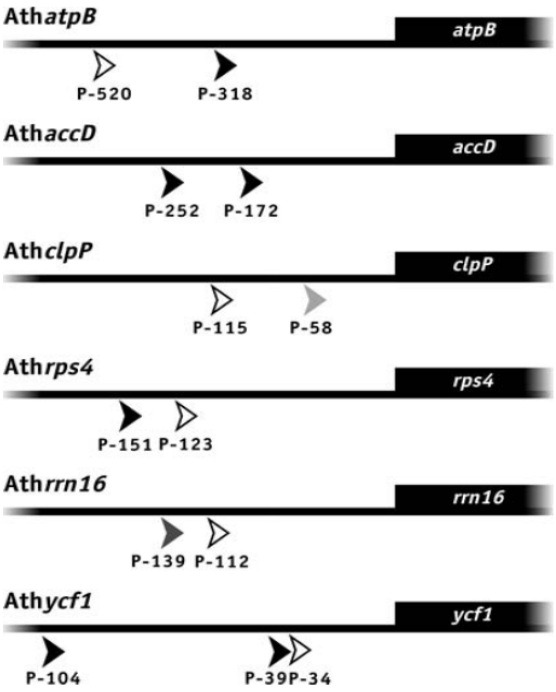


Fig. 4 Arabidopsis plastid genes with multiple promoters (Swiatecka- Hagenbruch et al., 2007). Schematic synopsis shows the multiple PEP and NEP promoters of Arabidopsis genes. Boxes represent genes, while open arrowheads denote PEP promoters, filled black arrowheads type-I NEP promoters, filled light gray arrowhead the PclpP-58 type-II NEP promoter, and filled dark gray arrowhead the Prrn16-139 Pc promoter. The promoters are named based on their position in respect to the translation initiation site (+1).

1.1.4.3 Specific functions and regulation of RpoT polymerases in plants

A. RpoTp

Both PEP and NEP are indispensable for the development of photo-synthetically active choloroplasts(Hess & Borner, 1999). Chloroplast differentiation is triggered by RpoT polymerases via activation of the chloroplast genetic system during early leaf development. NEP was shown to play a role at a specific stage of leaf development that is, prior to the

expression of PEP subunits and the photosynthetic apparatuses (Kusumi *et al.*, 1997; Kusumi *et al.*, 2000; Kusumi *et al.*, 2004). It was shown that the RpoTp enzyme level in maize leaves decreases as plastids mature, while production of mRNAs by RpoT increases (Cahoon *et al.*, 2004). Emanuel et al. (Emanuel *et al.*, 2006) observed that *RpoTp* is the most abundant of the three *RpoT* mRNAs in mature *Arabidopsis* leaves. Pronounced expression of *RpoTp* occurred in the green tissue of young leaves, and cell types rich in chloroplasts. However, *RpoTp* transcripts are also present in non-green roots. *RpoTmp* transcripts were found in green leaves. The *RpoTp* gene showed a delayed expression that, coincided with the development of photosynthetic competence, suggesting that RpoTp has a major role play in photo-synthetically active cells in green tissue. In contrast, the dual-targeted RpoTmp in plastids, functions principally in the transcription of genes in non-green types (Emanuel *et al.*, 2006).

RpoTp was found to be more selective than RpoTm in promoter recognition (Liere *et al.*, 2004). Over-expression of RpoTp in tobacco has provided evidence for distinct promoter specificity (Liere *et al.*, 2004). To acquire information on the individual role of RpoTp in plastid transcription, Swiatecka-Hagenbruch et al. (Swiatecka-Hagenbruch *et al.*, 2008) have studied transgenic *Arabidopsis* plants with both lowered activity of RpoTp and mutants lacking functional RpoTp. They showed differential effects of the altered RpoTp activities on the usage of NEP promoters, providing evidence for the importance of RpoTp for transcription of a subset of NEP promoters.

B. RpoTm

The amount of transcripts of all three *RpoT* genes (*RpoTm*, *RpoTp* and *RpoTmp*) was reported to be low in all plant organs, with flower tissues having the highest relative accumulation. *RpoTm* transcripts were the most abundant in all organs with the exception of mature leaves, where RpoTp transcripts showed the highest accumulation (Emanuel *et al.*, 2004). In *Arabidopsis*, both RpoTm and RpoTmp are highly expressed in meristematic and young tissues with high mitochondrial activity (Emanuel *et al.*, 2006). RpoTm can recognize mitochondrial promoters accurately *in vitro*, appearing to recognize sequences having a TATATA element. Conversely, RpoTmp did not display any significant promoter specificity (Kuhn *et al.*, 2007). By using novel mutations in the *A. thaliana RpoTm* gene, Tan et al. (Tan *et al.*, 2010) found that RpoTm is the major RNA polymerase in mitochondria. It can transcribe most, if not all, mitochondrial genes, which is sufficient for normal gametophyte and embryo development. Tan et al. (Tan *et al.*, 2010) showed that RpoTm is important for normal pollen tube growth, female gametogenesis and embryo development. It also has

distinct genetic and molecular roles in plant development, which cannot be replaced by RpoTmp.

C. RpoTmp

Mutations in the *RpoTmp* gene affected the light-induced accumulation of several plastid mRNAs and proteins and resulted in a lower photosynthetic efficiency (Baba *et al.*, 2004). In *Arabidopsis* mutations in RpoTmp reduced plant size, but did not affect male and female gametophyte and embryo development (Tan *et al.*, 2010). It was clearly demonstrated by Swiatecka-Hagenbruch et al. (Swiatecka-Hagenbruch *et al.*, 2008) that RpoTmp is involved in chloroplast transcription and might play a major role in transcription from *PclpP-53* (type-II) and *Prrn-139 (Pc)* NEP promoters. Furthermore, RpoTmp was shown to perform gene-specific transcription (*rps4*, *nad2*, *nad6*, *cox1* and *ccmC* genes) in mitochondria of *Arabidopsis thaliana*, thghlighting its participation in mitochondrial gene transcription (Kuhn *et al.*, 2007; Kuhn *et al.*, 2009). Major differences in transcript abundances were revealed in those studies between wild-type and RpoTmp mutants (see also 1.1.4.2).

1.1.5 Origin and evolution of RpoT polymerases

In addition to T-odd phages, *RpoT* sequences have also been found in linear mitochondrial plasmids (Cermakian *et al.*, 1997). Neither of the two can be considered as a likely source for the RpoT polymerase of the eukaryotic kingdom. The T-odd phages are lytic proteobacteria infecting viruses and the plasmid encoded enzymes were only found in some plant and fungal mitochondria and not all appear to encode functional enzymes (Cermakian *et al.*, 1997). A podoviral origin of the RpoT polymerases was proposed based on the observation (Filθe & Forterre, 2005; Shutt & Gray, 2006) that a prophage might have been part of the proteobacterial endosymbionts genome or that an active podoviral phage was present in the proteobacterial ancestor of mitochondria at the time of the endosymbiotic uptake. *RpoT* genes from angiosperms have been well characterized (Hess & Borner, 1999; Weihe, 2004), however, a lack of data on *RpoT* genes for green algae and lower plants leaves the early molecular evolution of the *RpoT* gene family unresolved.

Although the origin of the *RpoT* genes remains unclear (Cermakian *et al.*, 1997), the study of the *RpoT* genes and their gene products in a variety of higher organisms including a number of land plants has shed light onto the molecular evolution of this key component of organellar transcription. The information about the types and phylogenetic distribution of RNAPs, and their structural similarity to one another, is necessary to determine whether the RNAPs of different eukaryotes all arose from a single common ancestor and what the evolutionary

source of this enzyme might have been. Phylogenetic analyses (Richter *et al.*, 2002) show a distinct group of the plant *RpoT* sequences clearly separated from those of fungi and animals (Fig. 5). In concordance with this finding, the plant specific gene structure is highly conserved in all plant *RpoT* genes studied thus far (Richter *et al.*, 2002). In order to address the question of whether a T3/T7-like RNAP was acquired early or at a relatively late stage in the evolution of the mitochondrial transcription system it will be necessary to explore a phylogenetically broad range of earlier diverging unicellular eukaryotes (protists). Comparative information about RNAPs and other transcriptional components in different eukaryotes is also essential for defining and understanding species-specific peculiarities in the biochemical mechanism of expression of mitochondrial genomes that can vary tremendously in size, base composition and organization.

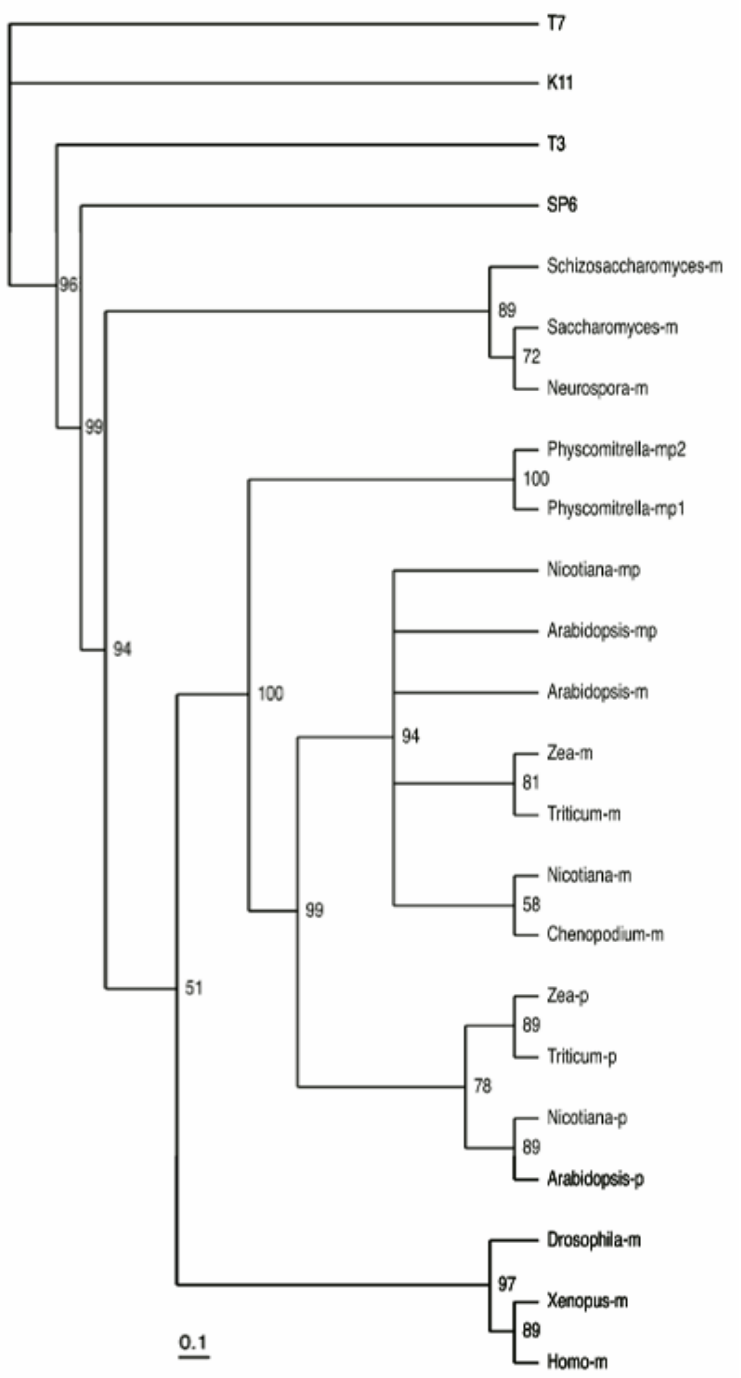


Fig. 5 Phylogenetic analyses of RpoT sequences (Richter et al., 2002). Sequences and their accession numbers (-m designates mitochondrial enzyme, -p plastid enzyme, -mp enzyme with dual targeting): T7, bacteriophage T7 RNAP (M38308); T3, bacteriophage T3 RNAP (X02981); K11, bacteriophage K11 RNAP (X53238); SP6, bacteriophage SP6 RNAP (Y00105); Saccharomyces-m, *Saccharomyces cerevisiae* (M17539); Neurospora-m, *Neurospora crassa* (L25087); Schizosaccharomyces-m, *Schizosaccharomyces pombe* (T38431); Homo-m, *Homo sapiens* (O00411); Xenopus-m, *Xenopus laevis* (AAF19376); Drosophila-m, *Drosophila melanogaster* (AAF51421); Arabidopsis-m, *Arabidopsis thaliana* RpoT1 (P92969); Arabidopsis-p, *Arabidopsis thaliana* RpoT3 (O24600); Arabidopsis-mp, *Arabidopsis thaliana* RpoT2 (CAC17120); Nicotiana-m, *Nicotiana sylvestris* RpoT1 (AJ416568); Nicotiana-p, *Nicotiana sylvestris* RpoT3 (AJ302020); Nicotiana-mp, *Nicotiana sylvestris* RpoT2 (AJ302019); Chenopodium-m, *Chenopodium album* (CAA69305); Zea-p, *Zea mays* RpoT1 (AAD22977); Zea-m, *Zea mays* RpoT2 (AAD22976); Triticum-m, *Triticum aestivum* RpoT-G (AAF32492);Triticum-p, *Triticum aestivum* RpoT-C (AAB01085); Physcomitrella-mp1, *Physcomitrella patens* RpoT1 (CAC95163); Physcomitrella-mp2, *Physcomitrella patens* RpoT2 (CAC95164).

1.2 Gene duplication and molecular evolution

Conservation of sequence and gene structure within the family of plant *RpoT* genes suggest that the molecular evolution of these genes was driven by gene duplication events. In recent years, the role of gene duplication in the emergence of novel gene functions has become one of the active topics of molecular evolution.

Gene duplication, i.e. the duplication of a DNA regions containing a gene, may take place as an error during homologous recombination, a retrotransposition event, or duplication of an entire chromosome (Zhang, 2003). Duplications result from the unequal crossing-over that occurs during meiosis between misaligned homologous chromosomes. The new copy of the gene is often free from selective pressure, meaning this kind of mutation has no deleterious effects to its host and therefore, leading to a faster mutation.

A major role for gene duplication in evolution was first proposed over 100 years ago (Taylor & Raes, 2004). However, it was not until the advent of genome sequencing in the late 1990s, that the prevalence and importance of gene duplication events were clearly illustrated. With the complete sequence of bacteria, archaebacteria and eukaryotes genomes, the numbers of duplicated genes for some organisms can be estimated as shown in Tab. 4.

Tab. 4: Prevalence of gene duplication

Species	Gene number	Percentage of duplicated genes (%)	References
Archaea			
Archaeoglobus fulgidus	2,346	30	Klenk et al., 1997.
Bacteria			
Helicobacter pylori	1,590	17	Tomb et al., 1997.
Haemophilus influenzae	1,709	17	Rubin et al., 2000.
Eukarya			
Saccharomyces cerevisiae	6,421	30	Rubin et al., 2000.
Caenorhabditis elegans	18,424	49	Rubin et al., 2000.
Drosophila melanogaster	13,601	41	Rubin et al., 2000.
Arabidopsis thaliana	25,498	65	The Arabidopsis Genome Initiative, 2000.
Homo sapiens	40,580	38	Li et al., 2001.

Sequence analysis, population genetic modeling and molecular techniques has led to rapid advances being made in discovering the mechanisms by which duplicate genes diverge in function and contribute to evolution. Gene duplication is believed to be the most critical evolutionary force since the emergence of the universal common ancestor. Genes resulting from gene duplication events are called paralogs and may code for proteins with different functions and/or structures. In contrast, genes created from a speciation event are called orthologous genes and code for proteins with similar functions in different species.

Gene duplication, with an average duplication rate of ~1% per gene per million years (Lynch & Conery, 2000) occurs individually, and, like a point mutation, can be fixed or lost among the population. Some paralogs will be muted and eventually eliminated following gene duplication, while retained paralogs may exhibit changes in DNA sequence or gene expression leading to sub- or neo-functionalization (Force *et al.*, 1999; Adams *et al.*, 2003; Wang *et al.*, 2004a). For those duplications fixed through the long evolutionary process, the long-term fate will still be determined by their functions (Zhang, 2003). Gene duplication generates redundant functions, as it is often not advantageous for the organism to have two identical genes.

One of the most important outcomes of gene duplication is the origin of novel function. However, many more duplications result in a related function rather than a completely new function. The most important contribution of gene duplication to evolution is providing new genetic material with somehow new gene functions. Without gene duplication, the genome plasticity of species in adapting to different environments would be severely limited. It is also hypothesized that gene duplication contributes to the evolution of gene networks (Wagner, 1994). Species-specific gene duplication can also lead to species-specific gene functions for species-specific adaptation (Zhang *et al.*, 2002), contributing to species divergence and origins of species-specific features. Although many duplicated genes might turn into pseudogenes, it is possible for some to acquire new functions. Identification of species-specific gene duplications will help to pinpoint the genetic basis of species' unique features.

Many gene families have experienced duplication during the evolution process, and gene copy can either remain similar or differ dramatically between organisms (Alvarez-Buylla *et al.*, 2000; Wang *et al.*, 2004b; Lin *et al.*, 2006; Zhou & Ma, 2008; Xu *et al.*, 2009), probably related to functional evolution. With regard to the *RpoT* genes, it is known that the single-cellular green alga *Chlamydomonas reinhardtii* has only one gene for a phage-type polymerases, that likely encodes the mitochondrial RNA polymerase. The situation might be

identical in other single cellular algae, but only limited information is available from genome data. From lower land plants, only the moss *Physcomitrella* has been investigated. Here the situation is different from *Chlamydomonas* and from higher land plants. *Physcomitrella* has three highly similar nuclear genes for phage-type polymerases like monocots, but two of them are dually targeted to mitochondria and plastids ((Richter *et al.*, 2002) and unpublished data). According to their position in phylogenetic trees, the gene duplication that led to these two *Physcomitrella RpoT* genes occurred independently of the duplication events that likely occured later during plant evolution. This resulted in the gene families now observed in monocots and dicots. However, the timing of the initial gene duplication event that gave rise to the plastid-targeted phage-type RNAP, and the occurence of a third copy to supply a dual-targeted RNAP as found in dicotyledonous plants is unknown.

1.3 Lower plant species for investigation of RpoT evolution

It is evident from the data above that duplications of *RpoT* genes for phage-type RNA polymerases have taken place several times during plant evolution. Since *Chlamydomonas* has only one such gene, gene duplication may have occurred not before the evolution of plants with differentiated tissues. Taxa for investigation of the evolution of phage-type polymerases in land plants were selected from major branches representing monophyletic clades of land plants, i.e. representatives of the *Selaginellales* and *Nymphaeaceae*. Data from these plants are expected to provide insights into the frequency of independent gene duplications among land plants. They will also allow us to determine if the first gene duplication that lead to the plastid phage-type polymerase seen in monocots and eudicots occurred before the evolution of angiosperms. As *Selaginellales* and *Nymphaeaceae* are proposed to have appeared early in vascular plants evolution, investigation into these two plants is expected to provide some evidence to clarify when the first and the second gene duplication (until now only known from the eudicots *Arabidopsis* and *Nicotiana*) occurred.

1.3.1 *Selaginella moellendorffii*

Selaginella is an enigma in the plant kingdom. The history of this genus can be traced back to hundreds of millions of years ago. Since the late Silurian/early Devonian when *Selaginella* appeared on the earth (Phillips TL & DiMichele WA, 1992), it has been unremarkable in appearance (shown in Fig. 6, from http://genome.jgi-psf.org/ Selmo1/Selmo1.home.html), never blossoming and without agronomic values. However, *Selaginella* has been a hotspot for plant researchers since the twentieth century (Williams S, 1937), due to its unique biology and importance in understanding the evolution of early vascular plants.

Fig. 6 ***Selaginella moellendorffii***

Selaginella is a lycophyte. Lycophytes constitute a monophyletic group of plants with microphylls and adaxial, reniform sporangia. The history of the lycophytes is longer than that of any other vascular land plants. Based on some extensive fossil records, lycophytes prevailed during the Devonian to mid-Carboniferous, reaching their acme approximately 310 million years ago (Phillips TL & DiMichele WA, 1992).

As shown in Fig. 7 (Banks, 2009), the lycophytes form a monophyletic sister clade to the other vascular plants and are phylogenetically placed between the bryophyte and fern/seed plant clades. Only three families of lycophytes survive today, i.e., the *Lycopodiaceae* (club mosses), the *Isoeteaceae* (quillworts), and the *Selaginellaceae* (spikemoss). Although they have long been distinguished as an ancient and distinctive group, the phylogenetic relationships between lycophytes and other land plants was not clearly determined until chloroplast, mitochondrial, and nuclear genes were compared (Qiu *et al.*, 2007; Raubeson & Jansen, 1992; Duff & Nickrent, 1999).

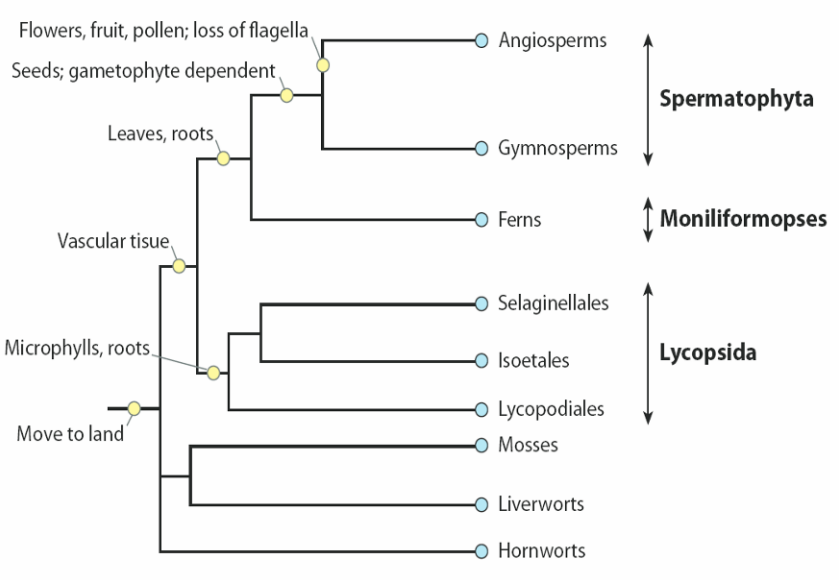


Fig. 7 Phylogenetic comparison of land plants, with some evolutionary features(The Green Plant BAC Library Project, http://www.greenbac.org).

The lycophytes that diverged from the euphyllophytes have retained some special characters of early land plants. They have a simple bifurcating apical meristem and a naked sporangium. Therefore, *Selaginella* is an ideal model from an evolutionary comparative perspective. *Selaginella moellendorffii* is a surviving member (400 million years) of an ancient vascular plant lineage without true leaves (having microphylls instead) and roots. Therefore, it represents an important node on the plant evolutionary tree and will act as a model organism to understand the evolution of plants as a whole, especially for comparative genomics.

1.3.2 *Nuphar advena*

Although most species of angiosperms are either monocots (especially grasses and orchids) or eudicots, the greatest diversity of structure (double fertilization, endosperm, "true" vessels, fibers, and great physiological and biochemical diversity) occurs in the basal angiosperms. Recent molecular phylogenies (http://www.greenbac.org/tree.html) have clearly identified 8-10 of the major basal flowering plant lineages that can represent this diversity in order to develop a comprehensive evolutionary picture of the ancestral angiosperm. Together with *Amborella, Liriodendron* and *Acorus*, *Nuphar* is one of the most representative basal angiosperms.

Based on comparisons of morphological features and interpretations of molecular analysis, the genus *Nuphar* is represented by eight distinct species in North America, including *Nuphar advena*.

Nuphar advena (waterlily, spatterdoc) is a second (*Amborella trichopoda* being the first) critical basal angiosperm. Distinct petals and sepals first appeared with the waterlilies, and they possess unusual "primitive" vessel types. This second basal branch of angiosperms, or perhaps the shared basal branch with *Amborella*, differs morphologically and anatomically from *Amborella* and thus is equally important for understanding the origin of key angiosperm innovations. *Nuphar* is becoming the focus of intensive developmental, reproductive, and evolutionary genomic studies, including floral biology and gene expression.

Nuphar advena is an aquatic perennial plant (see Fig. 8) that can be found in ponds or along slow moving water bodies. The habitats of the waterlily are characterized by shallow wetlands with a muddy bottom Sunlight is necessary to produce flowers.

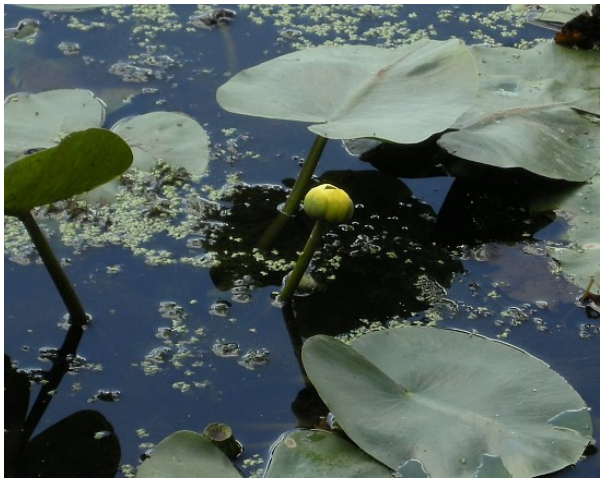


Fig. 8 Appearance of *Nuphar advena*

Its broad leaves are ovate or heart-shaped with smooth margins and blunt tips. Individual “half-opened” yellow flowers stand on thick stems a few centimeters above the water surface. Flowers are complete with sepals, petals, stamens and carpels (Skinner, 2006) and both male and female organs. Individual flowers are short-lived and replaced by ovoid fruits that are slightly constricted to the apex. The root system is rhizomatous which can produce vegetative colonies in shallow water.

2 Materials and Methods

2.1 Plants

Plants *Selaginella moellendorffii* and *Nuphar advena* were purchased from commercial suppliers, Dirk Wiederstein (Sessenbach, Germany, http://www.farn-gaertnerei.de/), and Seerosen Shop (Eschede, Germany, http://www.seerosensorten.de), respectively. The plants were grown in a growth chamber at 23 ºC with a light/dark regime of 8/16 hr. The intensity of light in all experiments was 210 µmol photons $s^{-1}m^{-2}$.

For *Arabidopsis thaliana* (ecotype Columbia), a seed stock from the Institute of Biology/Genetics (Humboldt University Berlin) was employed. Before plating, seeds were surface-sterilized by incubation in sterilization solution (for all buffer/solution composition, see 2.7) and subsequently washed with sterile water. Approximately 100-200 seeds were spreaded onto a Petri dish containing MS-agar medium. The dishes were sealed with Parafilm and then incubated for 4-5 days at 4-8 ºC for vernalization. The plates were then transferred to a growth chamber and kept at 23ºC under 8/16hr light/dark cycle. Twelve days later the young plants were transferred to soil and exposed to a light regime of 16/8hr light/dark.

2.2 Laboratory equipments and disposables

ABI 3130 automatic DNA Sequencer	Applied Biosystems, USA
Bio-Rad Universal Hood II System	Bio-Rad, USA
Corex tubes	Greiner bio-one, Frickenhausen, Germany
Counting chamber	Brand, Wertheim, Germany
Eppendorf tube	Eppendorf, Germany
Horizontal electrophoresis chambers	peqLab GmbH, Germany
Hybond-N^+-membrane	GE-Healthcare, Munich, Germany
Molecular Imager FX	Bio-Rad, Munich, Germany
Nylon net (60-80 µl)	Roth, Karlsruhe, Germany
Parafilm	Roth, Karlsruhe, Germany
Peltier Thermal Cycler PTC-200	Bio-Rad, California, USA
Petri dishes	Greiner bio-one, Frickenhausen, Germany
Phosphor imaging screen	Bio-Rad, Munich, Germany
QIAquick spin columns	Qiagen, Hilden, Germany
Radioactivity counter QC-2000	Bioscan,Paris, France
Razor blades	Roth, Karlsruhe, Germany
Spectrothermometer ND-1000	peqLab, Biotechnologie GmbH
Thin-walled PCR tube	Biozym, Oldendorf, Germany

2.3 Chemicals, Biochemicals, Biologicals

1 kb DNA ladder	Fermentas, St. Leon-Rot, Germany
100 bp DNA ladder	Fermentas, St. Leon-Rot, Germany
^{32}p-dCTP	PerkinElmer, Massachusetts, USA
Agarose	Biozym, Oldendorf, Germany
Alkline-Phosphatase	Fermentas, St. Leon-Rot, Germany
Ampicillin	Roth, Karlsruhe, Germany
Bacto Agar	Oxoid, Hampshire, England
Bacto Tryptone	BD, Heidelberg, Germany
ß-mercaptoethanol	Roth, Karlsruhe, Germany
BSA, fraction	Roth, Karlsruhe, Germany
$CaCl_2$	Roth, Karlsruhe, Germany
$Ca(NO_3)_2$	Serva, Heidelberg, Germany
Chloramphenicol	Roth, Karlsruhe, Germany
Chloroform	Roth, Karlsruhe, Germany
CTAB	Serva, Heidelberg, Germany
DanKlorix	Colgate-Palmolive, Hamberg, Germany
DNA loading buffer	Fermentas, St. Leon-Rot, Germany
E.coli competent cells (TOP 10)	Home-made, stock of Inst. of Biology
EDTA	Roth, Karlsruhe, Germany
Ethanol	Roth, Karlsruhe, Germany
Ethidium bromide	Roth, Karlsruhe, Germany
Formaldehyd	Roth, Karlsruhe, Germany
Formamid	Roth, Karlsruhe, Germany
Glycogen	Fermentas, St. Leon-Rot, Germany
HCl	Roth, Karlsruhe, Germany
IPTG	Roth, Karlsruhe, Germany
Isoamylalkohol	Roth, Karlsruhe, Germany
Iso-propanol	Roth, Karlsruhe, Germany
K-Acetat	Roth, Karlsruhe, Germany
KCl	Roth, Karlsruhe, Germany
Macerocyme R10	YAKULT, Tokyo, Japan
Mannitol	Roth, Karlsruhe, Germany
MES	Roth, Karlsruhe, Germany

$MgCl_2$	Roth, Karlsruhe, Germany
MOPS	Roth, Karlsruhe, Germany
MS	Roth, Karlsruhe, Germany
NaAc	Roth, Karlsruhe, Germany
NaCl	Roth, Karlsruhe, Germany
NaOH	Roth, Karlsruhe, Germany
NH_4-Acetate	Roth, Karlsruhe, Germany
N-lauryl-sarcosine-sodium	MP Biomedical, Illkirch, France
pDrive vector	Qiagen, Hilden, Germany
Phenol	Roth, Karlsruhe, Germany
PEG 4000	Fluka, Steinheim, Germany
Sucrose	Roth, Karlsruhe, Germany
SDS	Roth, Karlsruhe, Germany
SSC	Roth, Karlsruhe, Germany
Sulfadiazine	Roth, Karlsruhe, Germany
Tris	Roth, Karlsruhe, Germany
Trizol	Invitrogen, Karlsruhe, Germany
X-gal	Q-Biogene, Heidelberg, Germany
XL10-Gold ultra-competent cell	Stratagene, La Jolla, USA
Yeast extract	Serva, Heidelberg, Germany

2.4 Kits

ABI PRISM™ Big Dye Teminator Cycle Sequencing Ready Reaction kit	Applied Biosystems, USA
CapFishing Full-length cDNA Premix kit	Seegene, Rockville, USA
Concert Plant RNA Reagent	Invitrogen, Karlsruhe, Germany
Hexa lable DNA Labeling Kit	Fermentas, St. Leon-Rot, Germany
Phusion™ Hot Start High-Fidelity DNA Polymerase Kit	Biolabs, USA
QIAGEN PCR Cloning Kit	Qiagen, Hilden, Germany
QIAGEN plasmid mini kit	Qiagen, Hilden, Germany
QuantiTect reverse transcription kit	Qiagen, Hilden, Germany
RNA Cleanup kit	Qiagen, Hilden, Germany
SuperScript III RNase H- reverse transcriptase kit	Invitrogen, Karlsruhe, Germany

2.5 Enzymes

Alkline-Phosphatase	Fermentas, St. Leon-Rot, Germany
BamHI, restriction enzyme	Fermentas, St. Leon-Rot, Germany
BSA, fraction	Roth, Karlsruhe, Germany
BstXI, restriction enzyme	Fermentas, St. Leon-Rot, Germany
Cellulase R10	Serva, Heidelberg, Germany
DraI, restriction enzyme	Fermentas, St. Leon-Rot, Germany
EcoRI, restriction enzyme	Fermentas, St. Leon-Rot, Germany
HindIII, restriction enzyme	Fermentas, St. Leon-Rot, Germany
Klenow enzyme	Fermentas, St. Leon-Rot, Germany
Phusion hot-start DNA polymerase	Finnzyme, Espoo, Finland
RNase A	Serva, Heidelberg, Germany
RNase mix	Fermentas, St. Leon-Rot, Germany
RNase inhibitor	Fermentas, St. Leon-Rot, Germany
SalI, restriction enzyme	Fermentas, St. Leon-Rot, Germany
SmaI, restriction enzyme	Fermentas, St. Leon-Rot, Germany
SpeI, restriction enzyme	Fermentas, St. Leon-Rot, Germany
T4 DNA Ligase	Fermentas, St. Leon-Rot, Germany
Taq DNA polymerase	QIAGEN, Hilden, Germany
XbaI, restriction enzyme	Fermentas, St. Leon-Rot, Germany

2.6 Buffers and solutions

Ampicilline stock solution	100 mg/ml, dissolved in water or 96% ethanol
Chloramphenicolstock solution	34 mg/ml, dissolved in 96% ethanol
CTAB buffer	2% CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris, pH 8.0
EnzWash solution	0.4 M mannitol, 20 mM KCl, 20 mM MES, pH 5.7, 10 mM $CaCl_2$
Enzyme solution	1.3% (w/v) cellulase R10， 0.3% (w/v) macerozyme R10， 0.4 M mannitol， 20 mM KCl， 20 mM MES, pH 5.7. Heat at 55°C for 10min and cool it to room temperature then add 10 mM $CaCl_2$, 5 mM ß-mercaptoethanol and 0.1% (w/v) BSA.
LB medium	10 g Bactotyptone, 5 g Yeast extract, 10 g NaCl for 1 liter
10x ligation buffer	400 mM Tris-HCl, 100 mM MgCl2, 100 mM DTT, 5 mM ATP (pH 7.8 at 25°C).
10x MEN	0,2 M MOPS. 10 mM EDTA. 50 mM NaOAc pH 7,0
MMg solution	0.4 M mannitol, 15 mM $MgCl_2$, 4 mM MES, pH 5.7
MS-Agar medium	4.3 g MS salts, 30 g Saccharose, 0.5 g MES, fill up to 1 liter with water, adjust pH to 5.7 with KOH, autoclave after adding 8 g agar
10x MOPS buffer pH 7.0	200 mM MOPS, 50 mM NaOAc, 1 mM EDTA
P1 buffer	50 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 µg/µl RNase A
P2 buffer	200 mM NaOH, 1% SDS
P3 buffer	5M K-Acetat pH 5.0
PEG solution (40%, v/v)	4 g PEG 4000, 3 ml H_2O, 2.5 ml 0.8 M mannitol, 1 ml 1 M $CaCl_2$
Pre-hybridisation buffer	0.5 M Sodium phosphate pH 7.2 with 7% (w/v) H_3PO_4 and 7% SDS
RNA loading buffer (200 µl)	100 µl Formamide, 35 µl Formaldehyde, 10 µl 10x MOPS, 2 µl EtBr, 53 µl ddH_2O
SOC medium	20 g Bacto-tryptone, 5 g Bacto-yeast extract, 0.5 g NaCl, 2.5 ml 1M KCl, ddH_2O to 1000 ml. Adjusted to pH to 7.0 with 10 N NaOH, autoclaved and supplied with 20 ml of sterile 1 M glucose immediately before use
20x SSC	3 M NaCl， 0.3 M Tri-sodium citrate
Sterilization solution (62.5	20 ml DanKlorix, 40 ml ultrapure water, 2.5 ml

ml)	N-lauryl-sarcosine-sol (20%)
1x TAE	40 mM Tris; 20 mM acetic acid; 1 mM EDTA
50x TAE buffer	2 M Tris-HCl pH 8.0, 1 M sodium acetate, 20 mM EDTA
1x TE buffer	10 mM Tris, 1 mM EDTA
Transfer buffer for southern blot	0.4 M NaOH， 0.6 M NaCl
W5 solution	154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 2 mM MES, pH 5.7
Washing and incubation solution	0.5 M mannitol, 4 mM MES, pH 5.7, 20 mM KCl

2.7 Primers and other oligonucleotides

Tab. 5 Oligonucleotide primers for *Selaginella moellendorffii*

Primers	Sequences (5'-3')
cDNA primers	
SMin_p	GCG GAC ACA CGG TTC TAA ACA G
SMin_m1	GCT TGG GCC AAC GTT TCT G
SMin_m2	CGT GCC ACG TCA TTG CTC AC
Smgapp1	cac aca tca atg ttc cag taa gc
Smgapm1	TAT TAG GCT TGG GTG GCA AT
Smgapp2	AAT GGA GCA TAC CTG CAC CT
Smgapm2	TAT TAG GCT TGG GTG GCA AT
SMin_Pn	TGC TGA TGT TCT TGC GGT GAT
SM_gap2111_fw	CCG CGC CGT CAT GTG GAG
SM_gap2111_rev	GAT CGC CGC AAG AAC ATC AGC A
BAC/genomic primers	
SM_BAC_5end_2	AAG GAT TAC TTC CGA CCT GT
SM_BAC_5end	GGC GGC TTG GCA GGG TTT AG
SM_BAC_n3	CTT CCT CTG CGG TGC CC
SM_BAC_n2	TGG AAG ACC TCT CGA GGG AG
SM_BAC_n1	TTG TCT TTC AAG TAC TCT TTC T
SM_BAC_sel5	ACC CTT TCC AAG CTC GAG GAG G
SM_5R2_BAC1	AGG CTT CAC AGA TAC CCG AAA A
SM_BAC_del_p	TGG GTC GCA TAT GAT GAT GA
SM_BAC_del_m	GAC GCG CAC CGA GCC TCC AT
GFP constructs primers	
gfp-cons-1p	AGC GGA TCC GTC ATG TGG AGA GCA GCC G
gfp-cons-2m	AGC CCC GGG GCT GAA GAG CGG CAG TTG G
gfp-cons-3p	AGC GGA TCC GTA CGG GTA TTG CTG TAA ACC CTA A
gfp-cons-mut-p	GCG CCG TCA TCT GGA GAG C
gfp-cons-mut-m	GCT CTC CAG ATG ACG GCG C
RACE primers	
Sm3R1	ATG TTT GGC GAG GCA AGG
Sm3R2	CGC ATG AAC GTC ATA CTG AGG
Sm5R1	CCA CAA GGC CAC CTA TTC CAC
Sm5R2	GCT CCC AAG GTG TCG AGT G
Sm5R3	CGG ACC GCA GGT GCA G

Tab. 6 Oligonucleotide primers for *Nuphar advena*

Primers	Sequences (5'-3')
cDNA primers	
A-fw1	CCT TGG ATA CAC TTG GGA AT
A-rev1	CCG CAC CTA ACC ACC AC
A-fw5	AAG GTG GTG GTT AGG TGC GGA A
A-rev5	TTC CGC ACC TAA CCA CCA CCT T
Na-A-gap1p	TGC AAG CAG AAA CTG GCA CC
Na-A-gap1m	ATC AGT GAC CAT CTG CCG TCT G
Na-A-gap2p	GGG CTA GTG GAG GTC ACC TTG
Na-A-gap2m	TCC CAG GGC TGC ATA GTG TT
NaAgap3p1	TAC AAA TAA CCC AAG TGC TG
NaAgap3m1	CTT ATA CTG TGA CGG CAG AA
NaAgap3p2	AAA ACA ACC TTG ACT GCC TT
NaAgap3m2	CAG TAT CGG CTG CTC ATA GA
NaA-sp2p	GGG GAT TGT GCA AAG GTG ATT
NaA-sp2m	TAC TGT GAC GGC AGA AAT AGA A
C-fw1	ACT TCT TGC GGT GGT AGA GAG A
C-rev1	TCT CTC TAC CAC CGC AAG AAG T
C-fw2	GGT GTG ACG TGG AAC TTA AAC TAT
C-rev2	ATA GTT TAA GTT CCA CGT CAC ACC
C-fw3	CTG GCA AAT CTA TAT GCT GGC GGT G
C-rev3	TGC TAA ACG GCC ATC ATA
NaC-sp1p	GGC TGC CCT GGA GGA GAT G
NaC-sp1m	CTG TAA CAA AGG CTG GCT GGC T
NaP1p	CCC GTG CCC ACC ACC CTT C
NaP1m	GTG CCC ACC CTT GTC ATA CCC T
NaP2p	TTC CGG GAG GTC TCG TG
NaP2m	CAA TAG GTC TTT CTG CCG AAT C
NaP3p	ATG AAT GGG CAA GAC GAT GGG T
NaP3m1	CAC AAT GGC TTT AAT GCG GAT G
NaP3m2	TGC ACG AGC CGT CCA GGT GAA
NaP4p	TGA TTC GGC AGA AAG ACC TAT
NaP4m	TTA ATG CGG ATG CTA CAG ACA T
NaP10p	AAG CGT GCT ATC TTC TCC TCC
NaP10m	CGC CTC CGT CTC CAT CTT C
NaP11p	AAG TTG GGA AGC CGT CTC AT
NaP11m	TAG GTC TTT CTG CCG AAT CA
NaP12p	CAG TAA GAT GGA CCA CTC CT
NaP12m	CGG ATG CTA CAG ACA TGA T
NaP-sp1p	GTT ACA TTG TCT GCG TTG GAA G
NaP-sp1m	GCC ACC TTG AAA GTA ATC CC
RACE primers	
NaA-rev11	TCT TCT ATA CAA TCC GAC GCC TCT
NaA-rev12	AGC CCT TGA ATT TCT TCT ATC GC
NaA-rev13	AAA TCA CCA AAA TTC GGC TTG T
NaA-rev14	GCC AGT TTC TGC TTG CAC ATA TC
NaA-fw12	TGG GCT AGT GGA GGT CAC CTT
NaA-fw14	TTT TCG TGG ACG GGC ATA TC
NaA-fw15	GTT CAA GGT GGT GGT TAG GTG C
NaC-rev16	AGG GTA AAT TTG GTG CCA ACT T
NaC-rev17	GCC GAA GTA TCC TCA TCC GTC

NaC-rev18	TCA GAA CAC CCA TCC CTA AAC C
NaC-fw4	GAG AGA ATA TGG GCC AGT GGG
NaC-fw5	CCT TTA CCA GAG AAG CCA GAC A
NaC-fw6	GAT TTT CGA GGA CGG GCA TA
NaC-fw7	AAT CTA TAT GCT GGC GGT GTT G
NaP-5R2	GCC CAT TCA TCA ACA GCA CCA T
NaP-5R4	CGC TCC TGC ACG AGA CCT CC
NaP-5R5	AGC CAA GGT GGG TCG AGG AG
NaP-5R6	GGA GGG AAA TTG GAA GAA GGA A
NaP-5R8	CCT GGA GTG ATT CTG GAA GGG
NaP-3R1	CAG AAG ACT GCT TTT CCA CCA A
NaP-3R2	CAG GGG TGC ATG ATT CAT ATT
NaP-3R3	GCC CAT CCT TGA GAA CTT GTT A
Genomic & BAC primers	
Na_seq_Tm_1	CAG AGG TCT ATC CGC TGA AT
Na_seq_Tm_2	CTA GTG GGC GTC CTT CTG
NaAI-1p	GGG CTG ATG ATG ACT TCG AAT G
NaAI-1m	GCT TCG GCC TTA TTT GCA TTA G
NaAI-2p	CGG CAG GTG ACT AAG ATA GTG A
NaAI-2m	GTC TGA TCG AGC GGA GCT GAT A
NaAI-3p	ACG GCA GAT GGT CAC GGA TAT T
NaAI-3m	CAA GTC CTT TAC GAA CTA GAG G
NaAI-4p	TCG AGA CGA TAT GGT GTT ATT G
NaAI-4m	GAG GCA CCA ACA TTG GCA TGT A
NaAI-5p	GGT GAT TCC TTA CAT GCC AAT G
NaAI-5m	ACT GCT TCA CGC TGT TGT CTT G
NaAI-6p	GCG CAC ACA TGG AGC AAG ACA A
NaAI-6m	CAC AAA CTC GGC TAT AAG GTC G
NaAI-9m	TCT CTT CCC AGG GCT GCA TAG T
NaAI-10p	GGC TCT TGT AAT GGC TTG CAA C
NaAI-10m	TCA GCA ATT CCG GAG TAA ACA T
NaAI-11p	GAA CCT GGT TGC TGG GGA GAA G
NaAI-11m	GTG CAG CAC TTG GGT TAT TTG T
NaAI-12p	ATC CTA CAA ATA ACC CAA GTG C
NaAI-12m	GAA TGG CAT CCC GCT CCT TTA G
NaAI-13p	CTA AAG GAG CGG GAT GCC ATT C
NaAI-13m	ATC CCC CAG CCA GCT CAT AAT G
NaAI-14p	TGA CTG CCT TGG AGG AGA TGT T
NaAI-14m	AAG GGG AGT AGT CCA CCT CAC A
NaAI-15p	GTG AGG TGG ACT ACT CCC CTT G
NaAI-15m	TTT TAA CGC CAA AAC CTG AAG A
NaAI-16p	CTC TTC AGG TTT TGG CGT TAA A
NaAI-16m	ATC ATG TGA GAT CCA TCG AGG G
NaAI-17p	CCC AAA TTT TGT ACA TTC CCT C
NaAI-17m	TCG GCT GCT CAT AGA GTT CCA C
NaAI-18p	ACT GAG GGA GAA ATT TGT GGA A
NaAI-18m	GAT CCG GAA GAG GAG GAA AAC T
NaAI-10p5	CCA GCC GAG TCT GAC ACT
NaAI-10p6	GAC CTA GTC CAA CGG CTA CAA G
NaAI-11pm4	TTT AAT TTG GCT TTG GTT TAG G
NaAI-11pm5	GAG CTT TTT CTC CAA TCA TCC A
NaAI-5p3	TCA TAT GCG GAA AGG GGT TGT A
NaAI-5m2n	GCA GCG AGA GGT TTG GTA AAG T
NaAI-6p3	TCC AGT CCA CGG CAC CAC GAC

NaAI-10p2	TGC TTC CTA AAT ACC AAA ACC T
NaAI-11p2n	TGC TTA TAA TAT GCC AGA TTT G
NaAI-11m3	CAA AAT GGA CTC GGC TCA ACT T
NaAI-15m3	ATT TCT TAT CTA GCC CTC ACC A
NaAI-5p2	CAT TAT AAA TAT GGG GCA AAG C
NaAI-5m2	TCT GGG CTT GGC TCA GTC TCA T
NaAI-6p2	CCG ATA TGA TAA ACT CAG AAA A
NaAI-11p2	ATA TGC CAG ATT TGT CTC TAC A
NaAI-11m2	ATT TAC ACC GTT ACT AAA GCA G
NaAI-15m2	TGC AAT AGA AAT CCC GAA TGA T
NaAI-4m-p	GGC AAC TTG ACC CAG CAG AGT A
NaAI-5p3	TCA TAT GCG GAA AGG GGT TGT A
NaAI-6p3	ATC CAG TCC ACG GCA CCA CGA C
NaAI-seq7	CAC AAA CTC GGC TAT AAG GTC G
NaAI-10p2	CAT GGA CCA CCT GCT TCC TAA A
NaAI-11p-m	GCC CAT TTT AGT TAC ACA CTG A
NaAI-11p2	TCA GAT ATT GAC CTC AAA TCA T
NaAI-15m3	AAT TTC TTA TCT AGC CCT CAC C
NaAI-18m-p	GTG GGT GAG TTA GCT TCC ATC A
NaAI-5n1p1	TCA GAT GTA TGA GGG TCC GCA C
NaAI-10p6	GAC CTA GTC CAA CGG CTA CAA G
NaAI-11pm5	GAG CTT TTT CTC CAA TCA TCC A
NaAI-5n3p	CAG ATG TAT GAG GGT CCG CAC
NaAI-5n3m	GAG TCT GGG CTT GGC TCA GGT C
NaAI-5p5	GAA TGT ATT TAC CGC TGA TGG A
NaAI-10p4	TTA ACA CTT CGA TTA AAC ATG A
NaAI-11p-m3	ATC TGA ACT TTA TCC AAG ACC C
NaAI-1m3	AAG TAG ATT GGG CGA GAG GTT C
NaAI-5p4	CCA TAG AAG AGA GCA TAG AAA A
NaAI-6p4	TTT CCG TTT ACT GTT CGT GTT C
NaAI-10p3	GCT AAA GGC TCG CAC AAC CTA C
NaAI-11p-m2	AGG ACA AAA GGC AAA ACT AGA A
NaAI-11p3	GTG CAC CCT TGT ACA TAG ATA A
NaAI-15m4	CGC ATC TAT CTA CTG AAG TTT T
Na_seqA_seqII	CTT TAA TGC TTC GTG CTA ATG C
Na-seqA-seq3	GGT ATT AGA GCG GGA CAC AAG
Na-seqA-rev2	TTG GGT CAT TCT ATC CTT AAC
Na-seqA-seq4	ATT TGT GCC AAA AAC TCA AGG A
Na-seqA-rev3	CAT GCT GTT AGT TCT TCG TTG G
Na_seqA_seq5	GGC AAT TAT CAA AAT GTG ACC T
Na_A_seq5	CTT CTA ACA ATG GCA ATT ATC A
Na_A_revD	AAC TTG CCT TGA GAG AGT GGT C
Na_seqA_seq6	TGG CCA AAA TCA AGA CAT GAA T
Na_seqA_revE	GAT GCC ATG AAG GTA AAG AGG A
NaCI1p	TCA TGG TTG CGC TAG GGT TGT C
NaCI1m	GGC GCT GAA TCA TCC TTT CCT T
NaCI2p	CGG CAG GTG ACT AAG TTA GTG A
NaCI2m	ATA AGC CGT TTC AAT CAA CAA C
NaCI3p	GGA TGG TCC ACC TGA TAT TCG T
NaCI3m	TTT ATC AAG GCC CTT ACG AAC T
NaCI4p	ATG GTG TTA TTG AAT GTG ATC C
NaCI4m	GTG GCA CCA ACA TTG GCA TGT A
NaCI5p	GAT TCC TTA CAT GCC AAT GTT G
NaCI5m	GCT GTT GCC TTG ATC CAT GAG T

NaCI6p	GCA CTC ATG GAT CAA GGC AAC A
NaCI9m	TCT TCC CAG GGC TGC ATA ATG T
NaCI10p	GTT TGG GAA TTT GGG TTA TGC C
NaCI10m	AAC ATC TGC GGG TTT CTC ACC A
NaCI11m	ATC AAA AGC CTT GCA CGC ACA G
NaCI12p	ATC ATG CGA AGG GAT TCA CAA A
NaCI12m	GAA TGA CAT CCC GCT CCT TTA G
NaCI13p	CTA AAG GAG CGG GAT GTC ATT C
NaCI13m	ACA ATC ACC CAG CCA GCT CAT A
NaCI14p	CTG CCC TGG AGG AGA TGT TTC A
NaCI14m	CAA GGG GAG TGG TCC ACC TCA C
NaCI15p	GTG AGG TGG ACC ACT CCC CTT G
NaCI16m	TTT GCA GGC AAT AGC TGT CAT C
NaCI17p	ACT GCT TTT CCC CCT AAT TTT G
NaCI17m	AAT CGG CTG TTC ATA GAG TTC C
NaCI18p	ATG CGA TGT GGA TGA AAT GAA C
NaCI18m	ACG GTC TGG AAG AGG AGG AAA A
Na_seqC_seqI	CCA CTG GCC CAT ATT CTC TCT A
Na_BAC2_seq2m	ATA TCG GCT GAT CTT AAA TC
Na_BAC2_seq3m	GAA GGT TCA AGG TGG TGG TTA G
Na_BAC2_seq7	CCG CAC CCG AGT CAT GTC AC
Na_BAC2_rev2	GGC CGC ATC GAA TAT AAC TT
Na-BAC2-seq8	TCC TTA AGC TGG TAC AAG AAT A
Na-BAC2-seq9	GAG GGT GTT ATG AGC AAG ATG G
Na-BAC2-rev4	CAA TTA TGC TTT CTG GCC TGT A
Na_seqC_rev5	AGC AGC TTT ACG ACG GCG ACT C
Na_seqC_rev7	ATC CCG AAT TTG CTC CTC CAT C
Na_BAC2_seq10	GGT CTA GAT CCA TGT ATG GTA A
NaCI-14p2	GAA GGT TTA CAA TCA CGC AGA G
NaCI-11p2	ATA GAA ACC CAA TGT TAT CGT
NaCI-11m2	TAT GTT ATG TTG CCT CCT AAG
seqC-rev6m	AAG TTT GGG AAT TTG GGT TAT G
NaCI-1pm	GCA AAA GAT GGC AGA CCT CCT A
NaCI-11pm	CTA GTG ATT CAG ATG GCA AGT
NaCI-4p2	ATG TTT CGA CTC CAG CTT GAT G
NaCI-rev6	TGC AGA CAG GTT CCC AAA TAG A
NaCI-1pm	AGA GAA CTT GCC CAC CTC GTT T
NaCI-5mp	CAG CTA TGA CAA GGG TGC ACA T
NaCI-rev6	TCC GGA TCT GAC CTT TAC CAA C
NaCI-seq9	CAT TTG GAT CCT GGA CAT TGG
NaCI-11pm	GGA TGC AGT AGT GGC GTA GCT C
NaCI_11pm2	CAT TAA ACA AAC CAA ATC AGG T
NaCI_11p3	CAA TCC CTC ATT ATC ATT TCT T
NaCI_11m3	CCT TCA AAA GTT GTA CTA AAG A
NaCI_14p3	TGC ATC TTC CTT AGA TTT GGT T
NaCI-15pm	CCT CTG CCT CTT AAT CAT GAC C
NaCI_15pm2	GAT TTA TTC TTA AAC AGG ACC A
NaCI_18p2	CAA TAA CCA GAT TTT TCC CAA T
NaPI-1p	TGA ATG GGC AAG ACG ATG
NaPI-1m	AAT TTG GTT TTC CCT TGT GGA T
NaPI-2p	CAA AAA TCA GGA GGC TCT CA
NaPI-2m	ATT AAG TGG AGG TTG CAC ATA A
NaPI-3p	AAA CAG CTT ATG TGC AAC CTC C
NaPI-3m	CAA GCC CTT TAC GTA CAA GTG

NaPI-4p	TGA TCC ACT TGT ACG TAA AGG G
NaPI-4m	TTT GGT GGT ATC AAC ATC GG
NaPI-5p	CAT GCC GAT GTT GAT ACC AC
NaPI-5m	CGT TGT TGC TTT GCT CCG TG
NaPI-5m2	GTT GTT GCT TTG CTC CGT GA
NaPI-6p	TAT GTT ATG CGC ACT CAC GGA G
NaPI-6p2	ACA CCC GCC GCG CTT AAT GC
NaPI_6p3	TCA TGT AAC TCG CCT TGA TCG T
NaPI_6p4	CTC TAG GCC GCG ATT AAA TTC C
NaPI_6p5	ACC GGA TTC AGT CGT CAC TCA T
NaPI_6p6	CGG GTT GGA CTC AAG ACG ATA G
NaPI-6m	GCC CGC TTG CCC ATA AG
NaPI-7p	GGC AAG CGG GCG ACC TAT T
NaPI-7m	TGA CAC TCC ACC GCC ACC GT
NaPI-8p	GAC GGT GGC GGT GGA GTG TC
NaPI-8m	ATA GGC ACG CCC ACG AAA
NaPI-9p	TTG ATT CGG CAG AAA GAC CTA T
NaPI-9m	GCC CCA AAG CTG CAT AAT G
NaPI-10p	TCG TGC AAT GGT TTA CAG CAT
NaPI-10m	TCT CTT CTT GCA GCT ATA CCC
NaPI-11p	CGG CTT CAG TGA ACC TAG TT
NaPI-11m	AAG TTG GAT GCA CAG CAG GAT
NaPI-12p	CCA ACT TCT CTA CTT GCT CGT A
NaPI-12m	CAA GCT GCA CTA AAT AGG ATC T
NaPI-13p	AAA GAA GAT TGA AAG AGC GTG A
NaPI-13m	GCC AAG ACA TAA TAC TAC GTG C
NaPI-14p	GAG ATG TTC CAA GCT GCA CGT A
NaPI-14m	ACT GTT TCC AGT ATG GCT GCA C
NaPI-15p	TGG GCT TCC AGT TGT GCA G
NaPI-15m	TGC TCT CCC TCT GTA ATG CCA G
NaPI-16p	CCC TTC AGG TTC TGG CAT TAC
NaPI-16m	TAT GTG AAC CAT CAA GGG AGT G
NaPI-17p	CCA CCA AAC TTT GTT CAC TCC C
NaPI-17m	GGC ATG CTA TAA AGC TCA ACG A
NaPI-18p	TGA TGA TAT GAG CCG CAT TC
NaPI-18m	CGC CAC CTT GAA AGT AAT CCC
NaPI-18m2	TCA GAT CAA AAT CAC CAC GG
GFP constructs primers	
n-NaAgfp-1p	AGC TCT AGA ATG TGG AGG ATT GCC AAA AAA T
n-NaAgfp-2m	AGC GTC GAC GCG AAG ATG AAG CTA ACG ACG A
n-NaAgfp-3p	AGC TCT AGA CAT GGC TTC TAT CTC CTC TGG C
NaA_mut_p	ACG ATC TGG AGG ATT GCC
NaA_mut_m	CAA TCC TCC AGA TCG TTT T
n-NaCgfp-1p	AGC TCT AGA ATG TGG CGG TTT GCC AAA AAA
n-NaCgfp-2m	AGC GTC GAC GCA ACG ACG AAG CTA GCG A
n-NaCgfp-3p	AGC ACT AGT GGC CAT CTC CCT CAT CTC TCA
NaC_mut_p1	ATC CTG AAA ATT ATC TGG CGG TT
NaC_mut_m1	AAC CGC CAG ATA ATT TTC AGG AT
NaPgfp-1p	AGC TCT AGA CTG GCT TCA ACG GCC GCA GC
NaPgfp-2m	AGC GTC GAC GGT TGG TCT GAA TGC ATC TG
NaPgfp-3p	AGC TCT AGA ACC ATC TCC ATC TCT CCA CTT C
NaPgfp-mut-p	CCT TCC ACC ACG GCT TCA ACG
NaPgfp-mut-m	CGT TGA AGC CGT GGT GGA AGG

2.8 Isolation of nucleic acids

2.8.1 Isolation of genomic DNA from *Sm* and *Na*

Leaves were ground to fine powder in liquid nitrogen and incubated in three volumes of CTAB buffer containing 2% ß-mercaptoethanol for one hour with agitation at 60ºC. The lysate was extracted two times with chloroform/isoamylalcohol (24:1), and the nucleic acids were precipitated with isopropanol. The DNA pellet was washed with 70% ethanol and dissolved in TE buffer (10 mM Tris, 1 mM EDTA).

2.8.2 Isolation of BAC DNA

Bacterial stocks, containing the BAC clones, were plated to get single colonies on LB-agar plates containing chloramphenicol (12.5 µg/ml). Single colonies was picked and inoculated into 4 ml LB medium plus chloramphenicol (12.5 µg/ml), cultured at 37ºC and 220 rpm for about 5 hours. 1 ml culture was then transferred into 100 ml of LB medium with chloramphenicol (12.5 µg/ml) and incubated at 37ºC and 220 rpm for approximately 16 hours. BAC DNA was isolated using the QIAGEN plasmid midi kit according to the protocol of the manufacturer.

2.8.3 Small-scale plasmid DNA Isolation from *E. coli*

Three to four ml LB/Amp medium was inoculated with a single colony and incubated overnight at 37ºC and 220 rpm. Cultures were transferred into 2 ml Eppendorf tubes and cells were pelleted by centrifugation (14,000 rpm for 5 min at room temperature). Plasmid DNA was isolated from bacterial cells according to the manufacturer's protocol of the QIAGEN plasmid mini kit. DNA was eluted with 50 µl ultra-pure water.

2.8.4 Isolation of total RNA from Selaginella moellendorffii and Nuphar advena

RNA was extracted using the Concert Plant RNA Reagent (Invitrogen) and purified using the RNA Cleanup kit (Qiagen) according to the protocols of the manufacturers.

2.9 cDNA synthesis

cDNA was synthesized from 1 µg total RNA, isolated from leaves of *Selaginella moellendorffii* and *Nuphar advena*, respectively, using the QuantiTect reverse transcription kit, which eliminates genomic DNA contamination. Several cDNAs were prepared from 1 µg of total RNA using SuperScript III RNase H^- reverse transcriptase kit to produce longer cDNAs with higher yield. The reactions were either primed with Oligo-dT, random hexamer primers or sequence-specific oligonucleotides.

2.10 Determination of nucleic acid concentration

The quality and quantity of nucleic acids were visually examined in ethidium bromide-stained agarose gels. Additionally, UV absorption at 260 nm of the nucleic acid samples was measured using the spectrophotometer ND-1000 to determine the nucleic acid concentration. A ratio of A 260/A280 between 1.8 and 2.0 indicated a sufficient purity of the nucleic acids.

2.11 Agarose gel electrophoresis of DNA

DNA samples were separated by running agarose gels containing 0.8-1.5% (w/v) agarose and 0.2 µg/ml ethidium bromide in 1x TAE running buffer. Electrophoresis was carried out at 5-10 V/cm in a horizontal electrophoresis chamber. 1 kb or 100 bp DNA ladder was used as molecular weight markers. After separation, DNA molecules were visualized by UV transillumination employing a Bio-Rad Universal Hood II System (Bio-Rad). When required, DNA fragments were purified from excised gel pieces by QIAquick spin columns according to instructions of the manufacturer.

2.12 Agarose gel electrophoresis of RNA

1% (w/v) agarose gels for RNA analysis were prepared by melting agarose in ultra-pure water supplemented with 10x MEN (1/10 of the final gel volume) and formaldehyde (1/40 of the final gel volume) after cooling the agarose to 60 ºC. RNA samples were supplemented with 1.6 volumes of RNA loading buffer and incubated at 75 ºC for 10 min prior to loading. With 1x MEN as running buffer, electrophoresis was carried out in a horizonal electrophoresis chamber at 4 V/cm for 10 minutes and then changed to 8 V/cm. Separated RNA molecules were visualized by UV transillumination with a Bio-Rad Universal Hood II System (Bio-Rad).

2.13 PCR

Amplification of DNA fragments was performed in 50 µl reaction mixtures in thin-walled PCR tubes in a Peltier Thermal Cycler PTC-200. Primers for PCR were designed by Primer Analysis Software Oligo 6.9 and purchased from Operon Biotechnologies GmbH (Cologne). All used primers were listed in Tab. 5 and Tab. 6 (see 2.7. for primers and other oligonucleotides). DNA fragments were amplified from genomic DNA or cDNA using *Taq* DNA polymerase. Each reaction contained 2.5 units of *Taq* DNA polymerase, 10 pmol each of forward and reverse primer (see Tab. 5 and Tab. 6), 10 µmol of each dNTP, and 100 ng of *Seleginella* genomic DNA, or 200 ng of *Nuphar* genomic DNA, or an appropriate amount of cDNA, in 1 x PCR reaction buffer in a 50 µl reaction. Cycling conditions were as follows: 95 ºC/1min; 35 to 40 cycles of 95 ºC for 30 s to 1 min, 55-65 ºC (depending on the annealing temperature of the primers) for 30 s, 72 ºC for 1 min per 1 kb of amplicon length; 72 ºC for 5 min. PCR products were analyzed by agarose gel electrophoresis (see 2.11.). Phusion hot-start DNA polymerase (Finnzymes) was used for nested

PCR and long amplicon amplification. Reaction set-up and cycling conditions followed the protocol provided by the manufacturer.

2.14 Colony PCR

For PCR analysis of cloned DNA fragments directly from bacterial colonies, reactions were set up in 20 µl volumes containing 1 unit of *Taq* DNA polymerase, 4 pmol of each forward and reverse primer, 4 µmol of each dNTP, 2 µl of 10x reaction buffer and cells from a single bacterial colony. Cycling conditions were as follows: 95 ºC/1 min; 30 cycles of 95 ºC/30 s, 55 ºC/30 s, 72 ºC/1 min per 1 kb of amplicon length; 72 ºC/5 min. PCR products were analyzed by agarose gel electrophoresis (see 2.11.).

2.15 Cloning and sequencing of PCR products

For sequencing, PCR amplicons were cloned into pDrive vector (Qiagen) according to the instructions of the manufacturer. Chemically competent TOP10 *E. coli* cells were prepared according to a standard protocol (Sambrook J & Russell DW, 2001). Two µl of ligation mixture were mixed with the competent cells, which were incubated on ice for 30 minutes and then shifted to 42 ºC for 45 seconds and successively 2 minutes incubated on ice. After 45 min incubation at 37 ºC with 250 µl SOC medium, transformants were selected on solid LB medium containing appropriate antibiotics and X-Gal for blue/white selection via overnight incubation at 37 ºC.
Sequencing reactions were set up using PRISM^{TM} Big Dye Teminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) according to the manufacturer's instructions. Each reaction contained 300 ng plasmid DNA and 10 pmol primer (see Tab. 5 and Tab. 6), Cycle sequencing, product purification and product analysis on an ABI 3130 automatic DNA sequencer were carried out by Dr. M. Meixner (SMB GmbH, Berlin).

2.16 3'- and 5'- RACE reactions

3'- and 5'- RACE reactions were carried out using the CapFishing Full-length cDNA Premix kit (Seegene) together with Phusion hot start DNA polymerase according to the protocols of the manufacturers. For the cDNA synthesis, 2 µg of purified total RNA was reverse transcribed using the SuperScript^{TM} III RNase H- reverse transcriptase kit.
Gene-specific (gsp) primers designed for the RACE reactions were 22 – 30 nt long, with a GC content of 40 – 60 % and an annealing temperature of ≥ 65 ºC. Besides, nested primers were designed to overcome low specificity of the primary PCR reaction. 3'- and 5'- RACE adaptor primers were provided in the kit. Taking 5'- RACE as an example (shown in Fig. 9), three gsp primers pairing with the provided 5'-RACE adaptor primer, were designed based on the already known cDNA sequence. The first round of PCR was conducted using 5'-RACE adaptor primer and

gsp primer 5'R1, with the cDNA as the PCR template. The product of the first round of PCR was 1:100 diluted and served as the template for the second round of PCR using the 5'-RACE adaptor primer and gsp primer 5'R2. Similarly, the product of the second round of PCR was 1:100 diluted and served as the template for the third round of PCR using the 5'-RACE adaptor primer and gsp primer 5'R3. According to the primer binding sites to the cDNA, the product of the first round of PCR was expected to be 100-200 bp longer than that from the second round of PCR, which is again 100-200 bp longer than that from the third round of PCR.

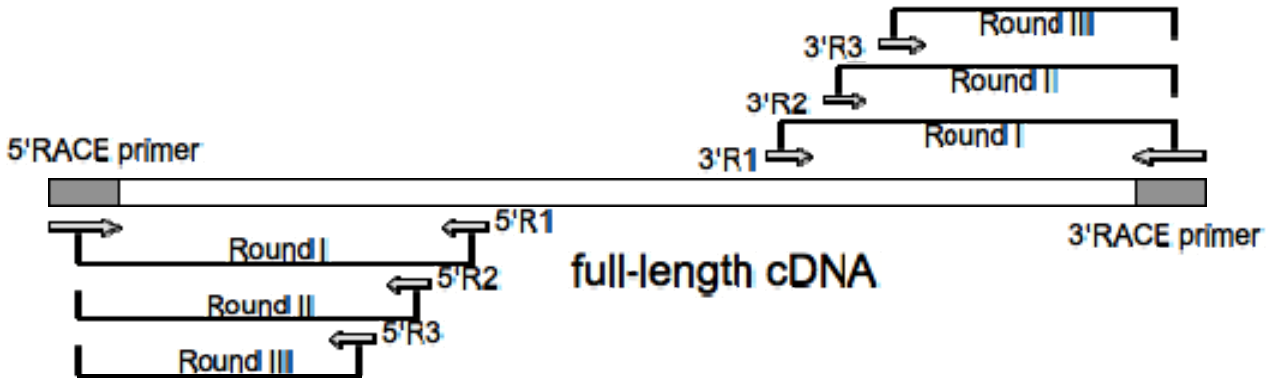


Fig. 9 Strategy of 3'- and 5'- RACE.

Phusion hot-start DNA polymerase was used to amplify the RACE PCR products. For a 50 µl reaction in thin-walled PCR tube, with the reaction mixture as listed in Tab. 7, the cycling conditions were as follows: initial denaturation at 98 ºC for 30 second, 35 cycles of denaturation at 98 ºC for 5-10 sec, annealing at 60-72 ºC for 10-30 sec and extension at 72 ºC for 15-30 sec/1 kb, followed by final extension at 72 ºC for 10 min.

Tab. 7 Reaction components of RACE-PCR

Components	Volume
5x Phusion HF Buffer	10 µl
10 mM dNTPs	1 µl
10 pmol primer 1	1 µl
10 pmol primer 2	1 µl
template	5 µl from 6.2.5.1
Phusion Hot Start DNA Polymerase	0.5 µl
ddH2O	add to 50 µl

2.17 Genomic cloning from BACs

2.17.1 BAC Filters

The *Selaginella moellendorffii* BAC library (*Selaginella*_HindIII BAC) was obtained from the GreenBAC Org (university of Arizona, USA) and had an average insert size of 123 kb (genome size: 132 Mb) covering 8.6 genome equivalents. It was constructed by Wang et al. (Wang *et al.*, 2005) at the HindIII site of pAGIBAC1 and contains 9,216 clones.

The *Nuphar advena* BAC library, also obtained from the GreenBAC Org had an average insert size of 135 kb (genome size: 3,040 Mb) covering 7 genome equivalents. It was constructed (Ammiraju *et al.*, 2006) at the HindIII site of pAGIBAC1 and contains 165,888 clones (on 9 filters). The format of the BAC filters and the positions of clones are shown in Fig. 10. Positive clones could be identified according to the different radioactive signal patterns after Southern hybridization (see Fig. 10D).

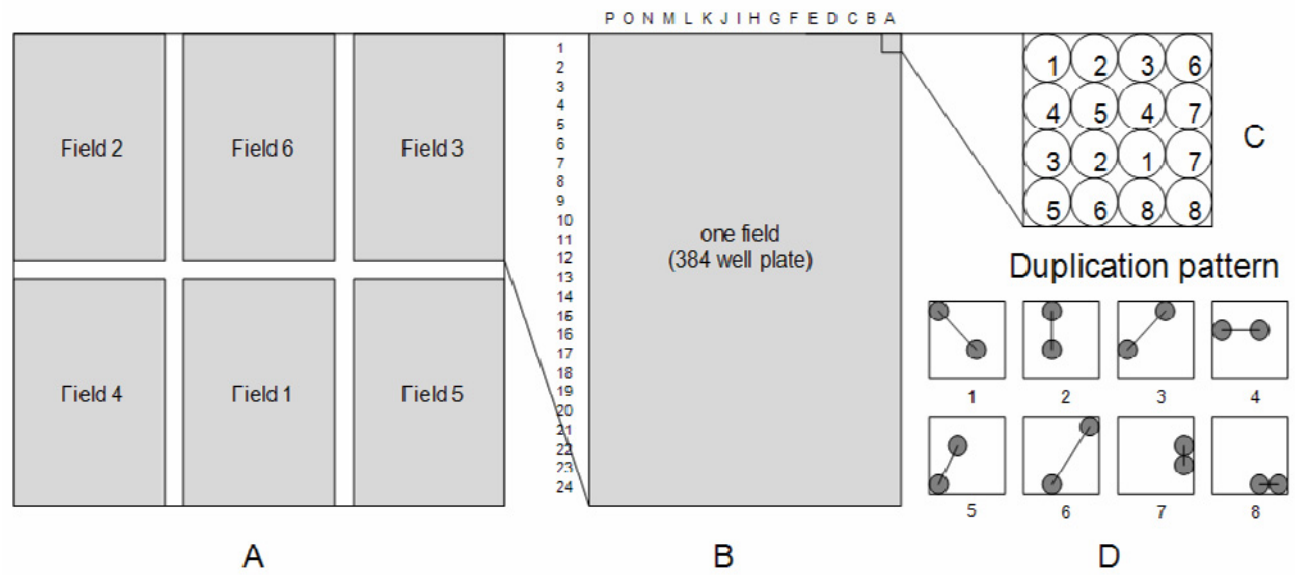


Fig. 10 The format of BAC filter and the positions of clones. A. Six fields of BAC clones on one filter. B. Each field corresponds to a 384-well format with 16x24 squares. C. For each square (e.g., enlarged well 1A), there are 16 (4x4) spotted positions for individual BAC clones. Each BAC DNA was spotted at characteristic positions two times to unambiguously score positives. In total, 36,864 clones (6 field x 24 rows x 16 columns x 16 positions per well) were spotted onto one BAC filter. D. Different patterns from southern hybridization according to the positions of different clones in C. Pattern 1-8 shows the positive radioactive signals from clone 1-8, respectively.

2.17.2 Hybridization of BAC Filters

2.17.2.1 Preparation of probe

A series of *Selaginella* trace sequence data was obtained from the trace database of NCBI (http://www.ncbi.nlm.nih.gov/Traces/) which showed high similarity to land plant *RpoT* sequences. PCR primers (see Tab. 5) were designed based on these sequences and used to obtain a 1.5 kb cDNA fragment by PCR reactions. PCR amplification was carried out in 50 µl reactions, using 1 x PCR buffer, dNTPs (200 mM each), primers (10 pmol for each), and Taq DNA polymerase (2.5 units per reaction) for 1 min at 95 °C, and 35 times for 30 s at 94 °C, 30 s at 63 °C and 90 s at 72 °C, followed by a final elongation step for 5 min at 72 °C. The 1.5 kb PCR product was labeled with ^{32}p-dCTP by using the Hexa label DNA Labeling Kit (Fermentas) according to the provided protocol. Briefly, 50-100 ng of the 1.5 kb PCR product was denatured at 98 °C for 10 minutes and annealed with random hexanucleotides on ice. Appropriate amounts of mix C, α-^{32}p-dCTP (50 µCi) and 5 units of Klenow fragment were added into the same 0.5 ml Eppendorf tube, followed by a DNA extension step at 37°C for 15 minutes, in order to allow the α-^{32}p-dCTP incorporate into the DNA molecules. The percentage of labeling was determined by measuring the specific radioactivity.

2.17.2.2 Filter hybridization

Both *Selaginella moellendorffii* and *Nuphar advena* BAC filters were screened using the 1.5 kb cDNA fragment as ^{32}p-labelled probe. Hybridization was performed in 7 % SDS, 250 mM sodium phosphate, pH 7.2 for 16-20 hours under non-stringent conditions (58 °C). Filters were washed with a series of solutions containing decreasing amounts of salts (2x to 0.5 x SSC, 0.1 % SDS) at 58 °C and exposed to a phosphor imaging screen, which was scanned in a Bio-Rad Molecular imager.

2.17.3 BAC DNA fingerprinting analysis

To discriminate different *RpoT* gene copies, southern hybridization (fingerprinting) of positive BAC clones was performed. Two µg of BAC DNA were digested with HindIII or BamHI in a 40 µl reaction volume, containing 4 µl of 10x buffer and 10 - 20 units of the restriction enzyme at 37 ^{o}C for overnight. The digested DNA was separated on a 0.8% agarose gel at 25 volt over 24 hours at 4 ^{o}C. The separated DNA was then transferred from the gel to an N^{+}-Nylon membrane(Sambrook J & Russell DW, 2001). After washing and drying, the Nylon membrane was applied to hybridization with a ^{32}P-labeled probe (for details of hybridization conditions, see 2.11).

2.18 Southern hybridization of genomic DNA

2.18.1 Restriction cleavage of genomic DNA

Based on the genomic sequence and specific cleavage sites, three restriction enzymes were chosen to digest the genomic DNA (BamHI, DraI, and BstXI for *Selaginella moellendorffii*, ScrFI, DraI and BseLI for *Nuphar advena*). The 40 µl reaction contained an appropriate amount of genomic DNA (10 µg for *Selaginella* and 20 µg for *Nuphar*) and enzymes (20 units for *Selaginella* and 30 units for *Nuphar*) in the specific reaction buffers. The reactions are listed in Tab. 8.

Tab. 8 Digestion reactions for genomic DNA from *Selaginella moellendorffii and Nuphar advena*

	Selaginella moellendorffii			Nuphar advena		
Restriction enzyme	BamHI	DraI	BstXI	ScrFI	DraI	BseLI
gDNA	10 µg in 30 µl			20 µg in 20 µl		
10x specific buffer	4 µl			4 µl		
Enzyme	2 µl			3 µl		
ddH_2O	4 µl			13 µl		
Total volume	40 µl			40 µl		
Temperature	37 °C	37 °C	55 °C	37 °C	37 °C	55 °C

2.18.2 Southern transfer and radioactive labeling of probe

The resulting fragments were separated by agarose gel electrophoresis (0.8% agarose gel, 30 volt, 4 °C, overnight). DNA was then denatured *in situ* and transferred from the gel to a N^+-Nylon membrane (Sambrook J & Russell DW, 2001). 100 ng of probe (see 2.17.2.1.) were radioactively labeled by random priming with Klenow enzyme and ^{32}P-dCTP. The Hexa label DNA Labeling Kit (Fermentas) was used according to the manufacturer's protocol.

2.18.3 Hybridization and image analysis

The Nylon membrane was hybridized under non-stringent conditions (58 °C) with the ^{32}p-labeled probe in 7 % SDS, 50 mM sodium phosphate buffer, pH 7.2, washed with a series of solutions containing decreasing amounts of salts (2x to 0.5 x SSC, 0.1 % SDS) . The filters were afterwards sealed in a plastic foil and exposed using a phosphor imaging screen. Radioactive signals were detected using a Molecular Imager FX detector. Images were directly imported into the software Quantity One (Bio-Rad, version 4.6.2) and analyzed.

2.19 Generation of GFP targeting constructs

2.19.1 pOL-GFP S65C plasmid

pOL-GFP S65C plasmid (Peeters *et al.*, 2000) is a modified vector based on pCK GFP-S65C (Reichel *et al.*, 1996) and pFF GFP-S65C with a small multi-cloning site. *Selaginella moellendorffii* and *Nuphar advena* fusion constructs were cloned into this vector via SpeI and SalI cutting sites. Fig. 11shows the GFP expression cassette of the pOL-GFP S65C vector.

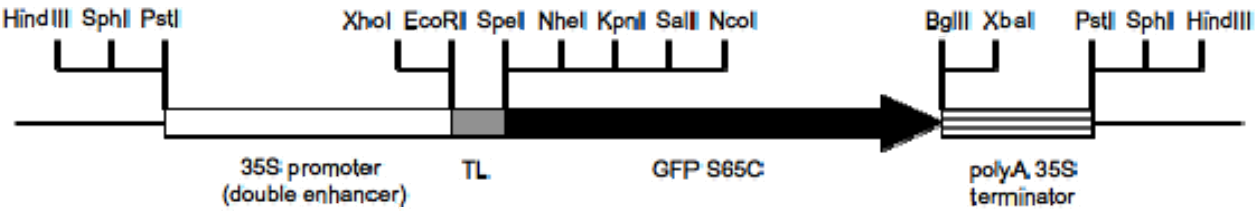


Fig. 11 GFP expression cassette of pOL GFPS65C. The unique sites of the multiple cloning sites (mcs) are in boldface. If no sequence is cloned in the mcs, the first ATG used is in the NcoI site and corresponds to the first methionine of the GFP protein. The vector contains a dual-enhancer 35S promoter from cauliflower mosaic virus (CaMV), the translation leader sequence from tobacco etch virus (TL), and the 35S polyadenylation signal from CaMV.

2.19.2 RecA-GFP and CoxIV-GFP

RecA-GFP (Renzette *et al.*, 2005) is a fusion of the amino-terminus of *Arabidopsis* recA and GFP which was used as a chloroplast (plastid) targeting control construct. CoxIV-GFP, containing the cox4 yeast transit peptide, was used as mitochondrial targeting control construct. It has already been shown that these two fusion proteins were faithfully targeted to each organelle upon expression in plant cells (Kohler *et al.*, 1997).

2.19.3 GFP targeting constructs

2.19.3.1 SmRpoT-GFP constructs

The amino-terminal sequences were amplified from cDNA of *S. moellendorffii* using the primers listed in Tab. 5. Products were ligated into vector pDRIVE and excised using XbaI and SalI. The fragments were inserted into pOL-GFP (Peeters *et al.*, 2000) which was opened with SpeI and SalI, to give the constructs shown in Fig. 12. Three constructs, namely, SmRpoT$_{met}$-GFP, SmRpoT$_{utr}$-GFP and SmRpoT$_{mut}$-GFP were established. The *SmRpoT* sequences (582 nt) for the full-length transit peptide GFP construct SmRpoT$_{utr}$-GFP was amplified using primers gfp-cons-3p and gfp-cons-2m. The shortened leader peptide sequences (276 nt), which started from the methionine initiation site, for construct SmRpoT$_{met}$-GFP, was generated using primers gfp-cons-1p and gfp-cons-2m. In the mutated construct, methionine (Met) was substituted by isoleucine (Ile)

using primers gfp-cons-3p and gfp-cons-mut-m, and accordingly primer gfp-cons-mut-p and gfp-cons-2m, to introduce the base change.

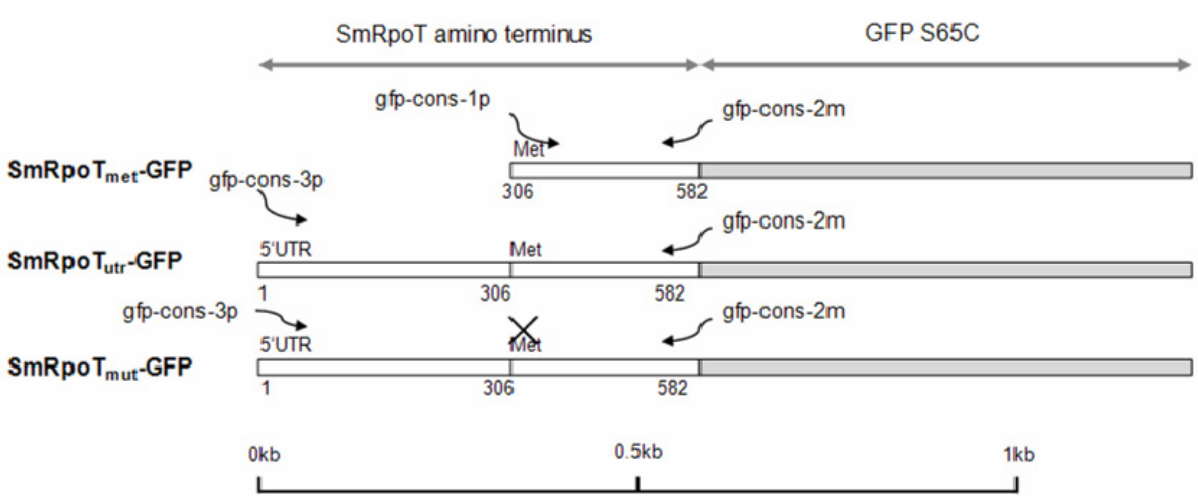


Fig. 12 SmRpoT-GFP fusion constructs. Amino-terminal RpoT sequences (white bars) were translationally fused to GFP S65C (gray bars) in plasmid pOL (see 2.19.1). The lengths of the fragments are given by nucleotide numbers (+1 is the 5' end of the 5'-UTR). The translation start is indicated by Met; the crossed Met position designates the mutation introduced at that position to prevent initiation of translation. The arrow pairs indicate the amplification directions. Primer pair 'gfp-con-1p' and 'gfp-con-2m' was used for SmRpoT-GFP construct, while primer pair 'gfp-con-3p' and 'gfp-con-2m' for both SmRpoTutr-GFP and SmRpoTmut-GFP constructs.

Primers gfp-cons-mut-p and gfp-cons-mut-m contained ATC or GAT (plus and minus strand, respectively) which substitutes the translation initiation codon ATG, resulting in amino acid mutation from Met to Ile (Fig. 13). In step 1, normal PCRs were carried out by using primer pair gfp-cons-mut-3p and gfp-cons-mut-m, and gfp-cons-mut-p and gfp-cons-mut-2m, respectively. Each reaction contained 2.5 units of *Taq* DNA polymerase, 10 pmol each of forward and reverse primer, 10 µmol of each dNTP, and 1 µl synthesed cDNA in 1 x PCR reaction buffer in a 50 µl reaction. Cycling conditions were 95ºC/1min; 35 cycles of 95ºC for 30 s, 62ºC for 30 s, 72ºC for 30 s; 72ºC for 5 min. For step 2, 1 µl product of each PCR reaction from step 1 was used as a combined template for PCR in a 50 µl reaction containing 5 µl 10x reaction buffer, 1 µl dNTP (10 mM each), 1 µl Taq enzyme and 41 µl ddH_2O. Cycling conditions were as follows: 95 ^oC for 1 min, 10 cycles of denaturation at 94^oC for 30 sec, annealing at 55^oC for 30 sec, and extension at 72^oC for 1 min, followed by additional extension at 72^oC for 5 min. Then, 10 pmoles each of primers gfp-cons-mut-m, and gfp-cons-mut-p were added into the reaction, and the PCR was continued under the same conditions for another 30 cycles (see Fig. 13).

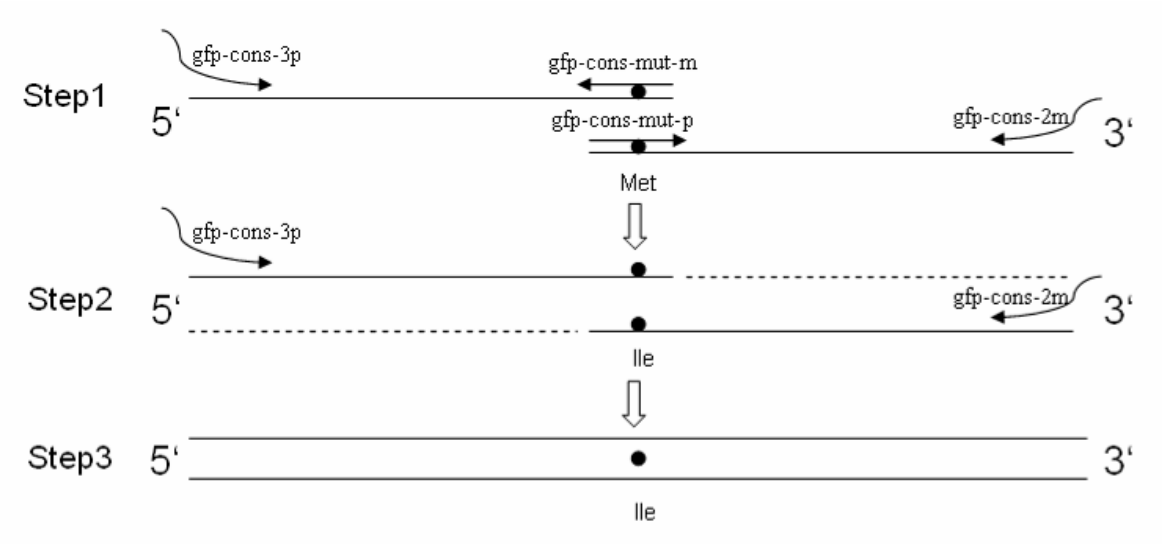


Fig. 13 Point mutation from Met to Ile for SmRpoT-GFP constructs.

2.19.3.2 NaRpoT-GFP constructs

Based on the method described above, three sets of NaRpoT-GFP constructs (Na-RpoTm1-GFP, Na-RpoTm2-GFP and Na-RpoTp-GFP) were established, as shown in Fig. 14, Fig. 15 and Fig. 16, respectively. The targeting sequences corresponding to the 118 N-terminal amino acids (353 nt) of construct Na-RpoTm1$_{met}$-GFP, 295 (884 nt) amino acids of Na-RpoTm1$_{utr}$-GFP and Na-RpoTm1$_{mut}$-GFP were PCR-amplified from *Nuphar advena* cDNA by using primers n-NaAgfp-1p, n-NaAgfp-2m, n-NaAgfp-3p, NaA_mut_p and NaA_mut_m, respectively (see Fig. 14).

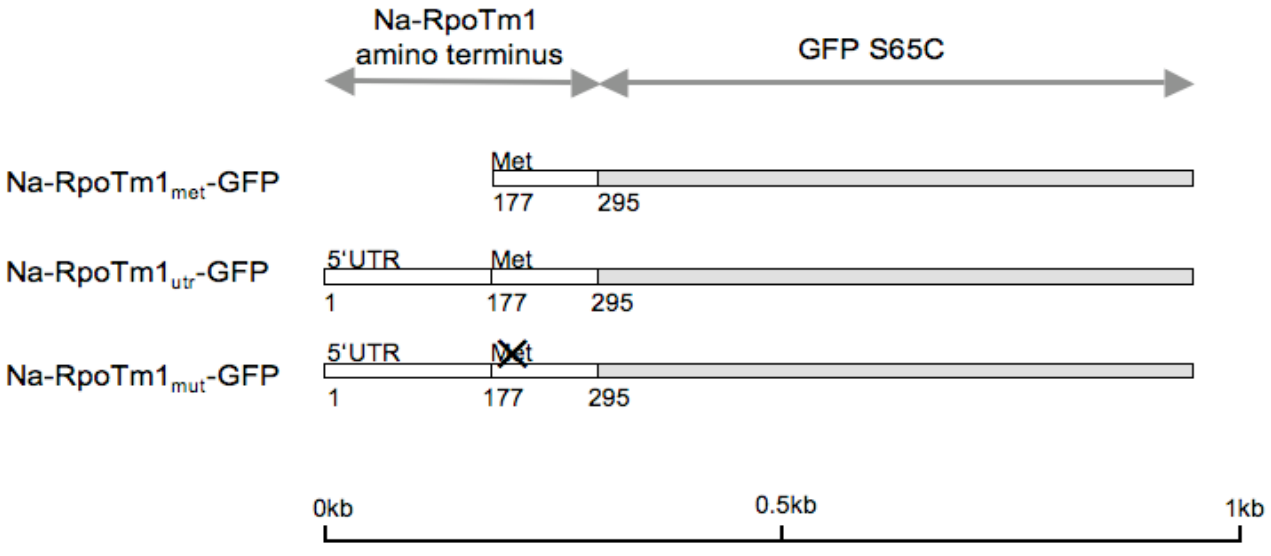


Fig. 14 Na-RpoTm1-GFP fusion constructs.

The targeting sequences corresponding to the 115 N-terminal amino acids (344 nt) of construct Na-RpoTm2$_{met}$-GFP, 368 amino acids (1107 nt) of Na-RpoTm2$_{utr}$-GFP and Na-RpoTm2$_{mut}$-GFP were PCR-amplified from *Nuphar advena* cDNA by using primers n-NaCgfp-1p, n-NaCgfp-2m, n-NaCgfp-3p, NaC_mut_p1 and NaC_mut_m1, respectively (see Fig. 15).

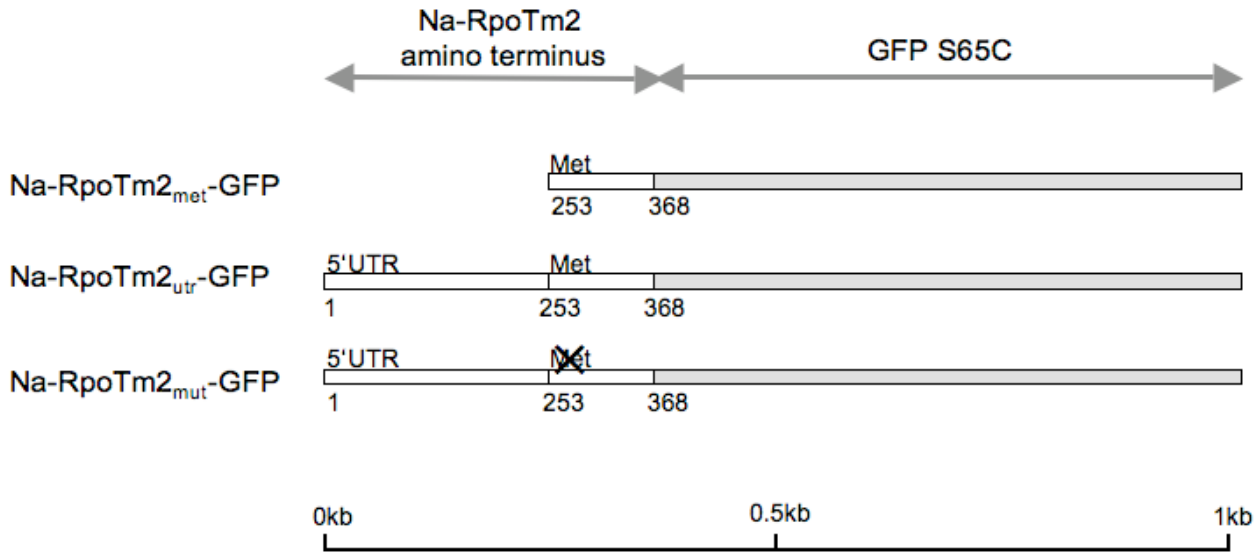


Fig. 15 Na-RpoTm2-GFP fusion constructs.

The construction principles for the three Na-RpoTp-GFP constructs (namely, Na-RpoTp$_{met}$-GFP, Na-RpoTp$_{utr}$-GFP and Na-RpoTp$_{mut}$-GFP, see Fig. 16) were the same as those for the other constructs. The only difference is that the initiation site is CTG (Leu), instead of ATG (Met) in the other GFP constructs, which was substituted by ACG (Thr) for construct Na-RpoTp$_{mut}$-GFP. PCR setup and reaction conditions were as described for the *Selaginella* constructs.

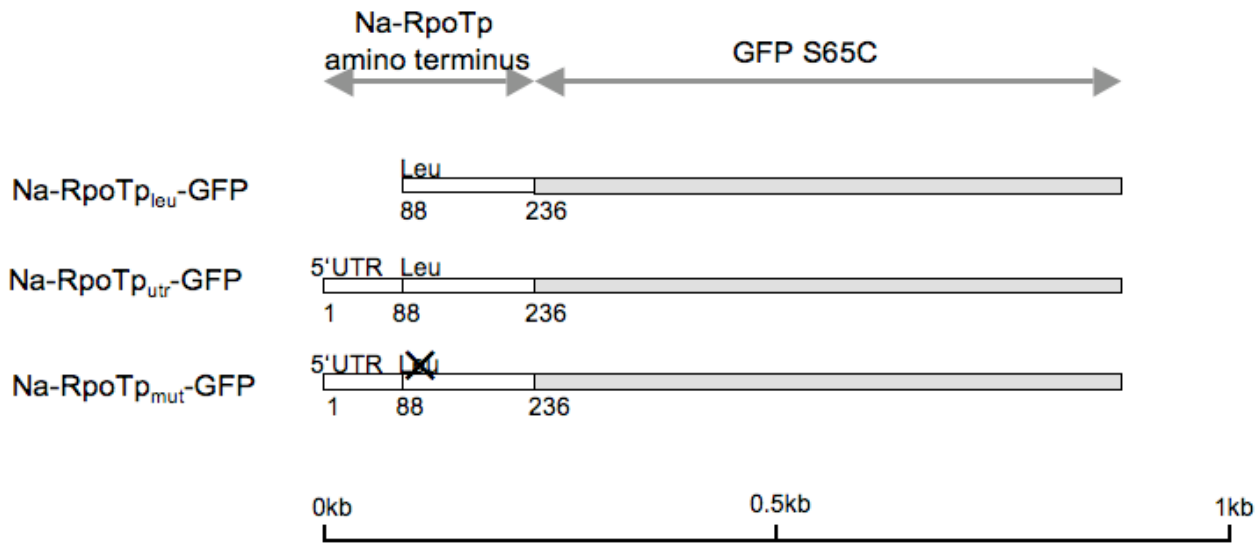


Fig. 16 Na-RpoTp-GFP fusion constructs.

2.20 Transient expression of GFP constructs

2.20.1 Isolation of protoplast from *Arabidopsis thaliana*

Protoplast was isolated from 3-5 weeks old Arabidopsis leaves grown under long day conditions (23 ºC, 16/8hr light/dark), as described by Sheen et al (Sheen's lab: http://genetics.mgh.harvard.edu/sheenweb/). Approximately 1 g of leaves was collected in a humid Petri dish. Leaves were cut in 0.5 - 1.0 mm parallel stripes with a razor blade on a clean glass plate. The stripes were immediately transferred with a scalpel into 15 ml EnzWash solution in a 9 cm glass Petri dish. The Petri dish was gently agitated (about 40 rpm, 1 min) and the EnzWash solution removed by pipetting. Then EnzSol solution was added, followed by gentle agitation at 40 rpm for 1min. Vacuum was applied for 30 min. The plant material was transferred onto a platform shaker and agitated at 40 rpm for 150 min in darkness. Agitation for 1 min at 80 rpm, followed by 30 min at 40 rpm in darkness, released the protoplasts. Finally, 10 ml solution W5 was added and the cells agitated for 1 min at 80 rpm. The cells were slowly poured through a 60-80 µm nylon net into a 30 ml Corex tube. The tube was centrifuged at 60 x g and 4 ºC in a swing-out rotor (HB-6) for 5 min. The supernatant was completely removed with a pointed glass pipette. The protoplasts were gently re-suspended in 15 ml W5 solution and centrifuged at 50 x g and 4 ºC for 5 min. The supernatant was removed as described above and the protoplasts re-suspended in 15 ml W5 solution. After 30 min incubation on ice they were centrifuged at 40 x g and 4 ºC for 5 min, and the supernatant was removed. Finally, the protoplasts were re-suspended in 0.5 - 1 ml MMg solution and counted in a counting chamber (0.2 mm depth) under the microscope. The protoplast concentration was adjusted to 2 x 10^5 /ml with MMg solution.

2.20.2 Isolation of protoplast from *Selaginella moellendorffii*

The procedure of *Selaginella* protoplast isolation followed the same steps as described above and was adopted from those described for maize (http://genetics.lgh.harvard.edu/sheenweb) and *Arabidopsis* (Yoo *et al.*, 2007) with several modifications: very young leaves were used for protoplasts isolation; leaves were collected and kept in distilled water for one hour for easy cutting into fine pieces; the enzyme solution contained 1.5% cellulase R10, 0.75% macerozyme R10, 0.6 M mannitol and 10 mM MES (pH5.7); a speed of 50 rpm agitation was applied, instead of 40 rpm for protoplast isolation from *Arabidopsis thaliana*.

2.20.3 PEG transformation

Ten µl DNA (10-20 µg of plasmid DNA - GFP fusion constructs) were mixed with 100 µl protoplasts in a 2 ml Eppendorf tube. 110 µl of PEG solution were added followed by thorough mixing. The cells were incubated at 23 ºC for 15 min and then diluted with 440 µl W5 solution. After spinning at 2000 rpm in a Biofuge at 23 ºC for 1 min, the supernatant was removed and the protoplasts were gently re-suspended in 100 µl, W5 solution, transferred to 24- well plates and 900 µl W5 solution was added to the cell suspension. The protoplasts were incubated at room temperature for 36-48 hours in darkness. For cell harvest, the protoplasts were transferred into a 2 ml Eppendorf tube and centrifuged for 1 min at 2000 rpm under room temperature. Most of the supernatant was removed, leaving the transformed protoplasts in approximately 30 µl.

2.20.4 Confocal microscopy

Transformed protoplasts were examined two days post transfection by confocal laser scanning microscopy with a Leica TCS SP2 using 488 nm excitation and two-channel measurement of emission from 510 to 580 nm (green/GFP) and > 590 nm (red/chlorophyll). Images were saved separately, with the green and red channels merged.

2.21 Computational phylogenetic analysis

Deduced amino acid sequences were aligned using ClustalW (Larkin *et al.*, 2007). Conserved blocks were cut out and merged as described earlier (Richter *et al.*, 2002) and subjected to Bayesian, maximum-likelihood and maximum parsimony analysis using programs Mr. Bayes 3.1 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) and Tree-Puzzle (Schmidt *et al.*, 2002), respectively. The Bayesian inference method employed the WAG amino acid replacement model with a gamma distribution to represent among-site rate heterogeneity (WAG +γ). MCMC was performed with 1 million generations and four independent chains. The Markov chain was sampled every 100 generations. Convergence was observed by plots of maximum likelihood (ML) scores and by using the run statistics. The first 10000 trees were discarded; the remaining trees were

used to construct a consensus tree and to calculate the posterior branch support values (alignment and Mr. Bayes' tree performed by Uwe Richter). In addition, maximum likelihood analysis with 100000 puzzling steps and maximum parsimony analysis with 100000 bootstrap replicates were conducted by program Tree-Puzzle.

3 Results

3.1 Identification of *RpoT* genes

Fig. 17 shows the general strategy for identification and isolation of *RpoT* cDNA (Fig. 17A) and genomic DNA (Fig. 17B) from *Selaginella moellendorffii* and *Nuphar advena*. Trace sequence data of *RpoT* sequences from *Selaginella moellendorffii* and *Nuphar advena*, available from the trace database of NCBI (http://www.ncbi.nlm.nih.gov/Traces/), were aligned with the *RpoT* gene sequences from other plant species. PCR primers were designed based on the multiple alignments and used to obtain partial cDNA sequences by RT-PCR. Afterwards, 3'- and 5'- RACE primers were designed and applied to amplify the 3' and 5' sequences. The full cDNA sequences could be obtained by primer walking.

As for the genomic DNA sequences, partial *RpoT* cDNA sequences were used as probes for hybridization with commercially obtained BAC filters (from GreenBAC, see "Material and Methods"). Positive clones were identified and purchased for BAC DNA isolation. Sequencing primers were designed based on the cDNA sequences to sequence the BAC DNA or sub-cloned fragments, respectively, to obtain the genomic DNA sequences for both *Selaginella moellendorffii* and *Nuphar advena.*

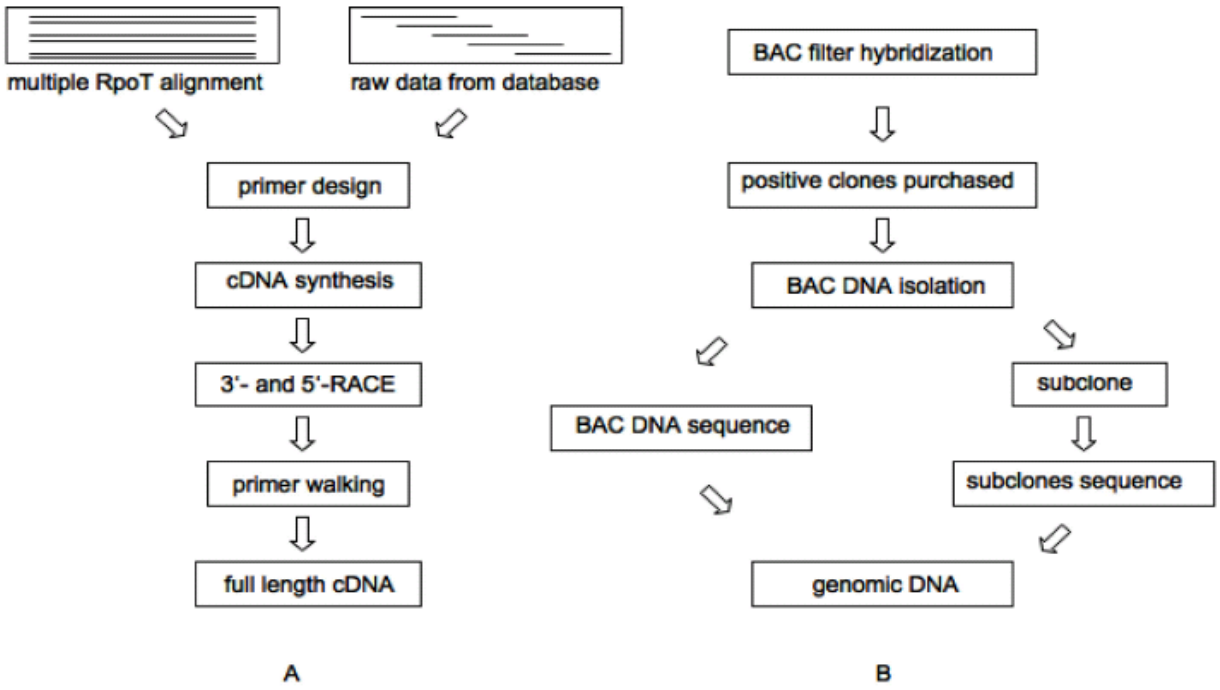


Fig. 17 The strategy to obtain full length cDNA (A) and gDNA (B) sequences.

3.1.1 Molecular cloning of the *RpoT* cDNA and gene of *Selaginella moellendorffii*

3.1.1.1 Identification of the full length *SmRpoT* cDNA sequence

Trace sequence data of *Selaginella* were found in the trace database of NCBI (http://www.ncbi.nlm.nih.gov/Traces/) which showed high similarity with the conserved region of angiosperm *RpoT* sequences. Based on these aligned raw sequences, gene-specific PCR primers were designed to identify the full-length *SmRpoT* cDNA sequence. Primers Sm3R1, Sm3R2, Sm5R1, Sm5R2 and Sm5R3 (see Tab. 5) were used to perform 3'- and 5'-RACE according to the strategy of 3'- and 5'- RACE shown in Fig. 9, to obtain the 3'- part and 5'-parts of cDNA sequences of *SmRpoT*. For details about the RACE amplifications, as well as the localization of the primers within the cDNA sequence, see "**Material and Methods**".

Fig. 18 shows the results of 5'- and 3'-RACE. Fragments from the three lanes (1, 2 and 3) as shown in Fig. 18A, were obtained by three rounds of PCR with three primer pairs, i.e., Sm5R1 and 5'-RACE adaptor primer, Sm5R2 and 5'-RACE adaptor primer (PCR products from the first round of PCR were 1:100 diluted and used as template), and Sm5R3 and 5'-RACE adaptor primer (PCR products from the second round of PCR were 1:100 diluted and used as template), respectively. Whereas after the first round of PCR only faint bands were observed on the agarose gel, the second and third round resulted in clearly visible DNA fragments of expected lengths (see Fig. 18A). The same principle was applied for the 3'-RACE reaction (see Fig. 18B). Primers SMin_p, SMin_m1, SMin_m2, Smgapp1, Smgapm1, Smgapp2, Smgapm2, SMin_Pn, SM_gap2111_fw and SM_gap2111_rev (see Tab. 5) were used to amplify the full-length cDNA sequence. The obtained fragments were purified and subjected to sequencing and then assembled using SeqMan software (DNASTAR, version 7.2.1).

The full-length *SmRpoT* cDNA is 3,452 bp long and constitutes 373 bp of 5' untranslated leader, 3,006 bp of protein coding sequence (CDS), and 73 bp of 3' untranslated sequence (Tab. 9), with the molecular map shown in Fig. 19. The sequence of *Sm RpoT* gene was deposited in the EMBL database under the accession number AM933560.1 (for sequences, see Appendix).

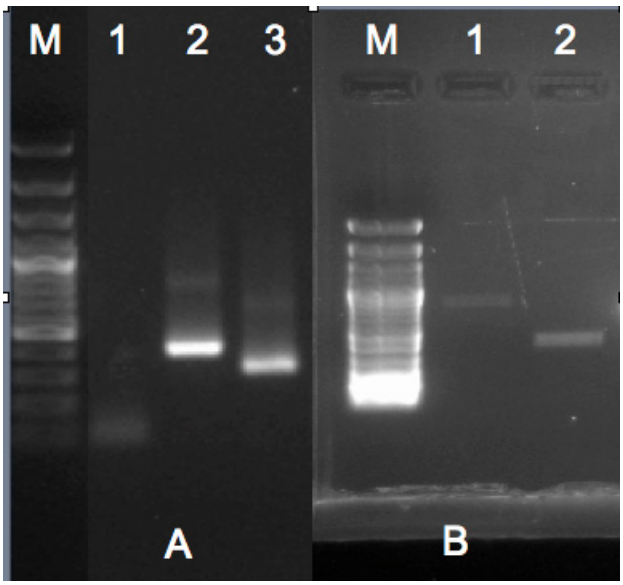


Fig. 18 Detection of 5'- and 3'-RACE products of *Sellaginella moellendorffii*. A. PCR products of 5'-RACE from three rounds are shown in lanes 1, 2 and 3. B. PCR products of 3'-RACE from two rounds are shown in two lanes 1 and 2. M: marker, 100 bp ladder plus.

Tab. 9: Characteristics of full-length cDNA of *SmRpoT*

Gene length	4,551 bp
Full-length cDNA length	3,452 bp
CDS	3,006 bp
Encoded amino acids	1002 Aa
5' UTR	373 bp
3' UTR	73 bp

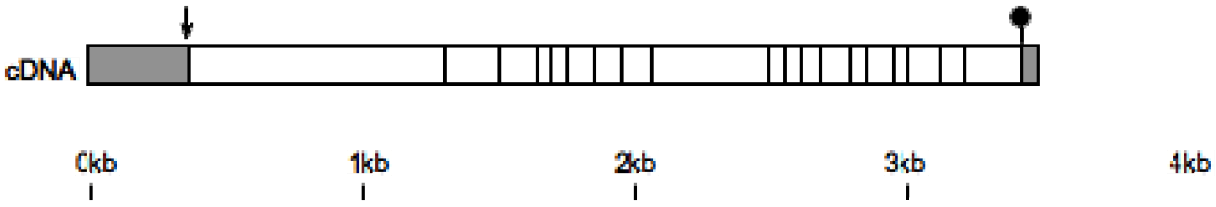


Fig. 19 Molecular map of SmRpoT cDNA. The translation initiation site (AUG) is designated by an arrow and the stop codon by a black circle. The 5' and 3' untranslated regions are designated by hatched areas. Vertical bars designate the position of introns.

3.1.1.2 Isolation of the genomic DNA sequence of *SmRpoT*

To elucidate the gene structure of the *RpoT* gene of *Selaginella moellendorffii*, the genomic sequence was isolated from a *Selaginella moellendorffii* BAC library obtained from the Green BAC Project (Arizona Genomics Institute). The BAC Filter was hybridized under non-stringent conditions with a 1.5 kb probe which was amplified from *Selaginella* cDNA as described in section

2.17.2.1. Five positive BAC clones were identified as shown in Fig. 20 and purchased for further analysis, namely, SM_Ba 0003O16, SM_Ba 0007B18, SM_Ba 0008A17, SM_Ba 0008B23 and SM_Ba 0011L8.

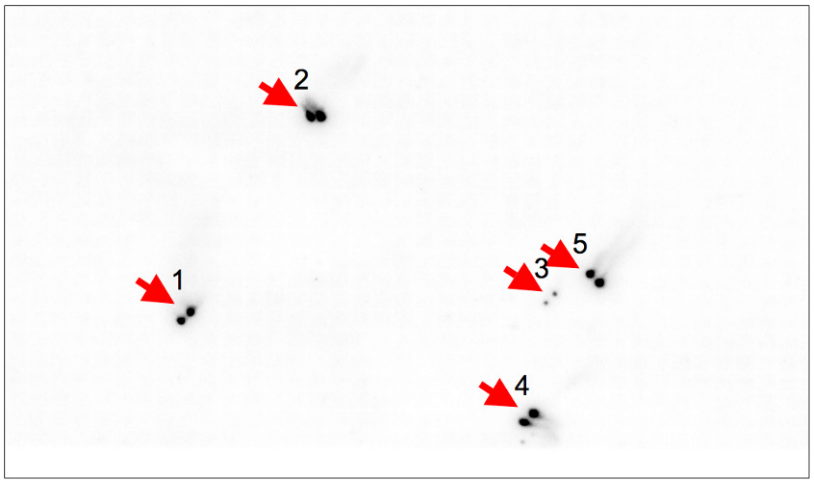


Fig. 20 Hybridization of the Selaginella BAC filter with a *RpoT* cDNA probe. Five positive clones were identified: Clone 1, SM_Ba 0007B18; clone 2, SM_Ba 0011L8; clone 3, SM_Ba 0008A17; clone 4, SM_Ba 0008B23; clone 5, SM_Ba 0003O16.

The five BAC DNAs were purified (Fig. 21A) and digested with enzymes HindIII and BamHI (Fig. 21B), respectively. After electrophoretic separation, the fragments were transferred to Nylon membrane and hybridized with the same probe used for BAC filter hybridization, under the same hybridization conditions. As shown in Fig. 21C, BAC clones 1, 2 and 5 exhibited the same fingerprinting pattern in both the HindIII and BamHI digestion, respectively. No hybridization signals were observed in BAC clones 3 and 4, which were therefore considered to be false positives or misidentified from the BAC filter hybridization. The fingerprint analysis showed that three (clones 1, 2 and 5) out of the five clones, originally identified, had identical HindIII and BamHI patterns after hybridization, and thus should originate from the same genomic region. This result also suggests that there is only one copy of an *RpoT* gene in *Selaginella* (for Southern hybridization analysis of gene copy number, see below, section 3.3).

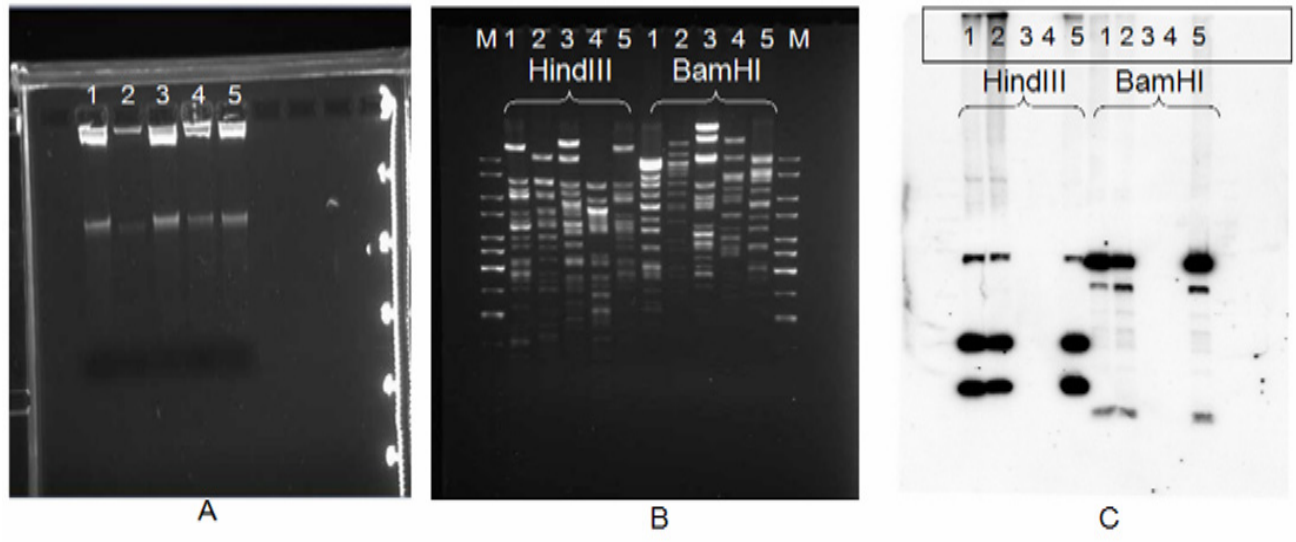


Fig. 21 Isolation and fingerprinting analysis of five *RpoT*-specific BAC clones from *Selaginella moellendorffii*. A. Purified BAC DNA of the five positive clones. B. Digestion with HindIII and BamHI, respectively. C. Hybridization with *RpoT*-specific probe. M, marker, 100 bp ladder plus. (See text for details).

The *SmRpoT* genomic DNA was obtained from BAC clone SM_Ba 0003O16 (clone 5) by using cDNA derived primers (see Tab. 5) for subsequent primer walking, directly from purified BAC DNA. Sequencing showed that the *SmRpoT* gene is 4.5 kb in length, and comparison of the genomic DNA with the cDNA defined 19 exons and 18 introns (see Fig. 22 and Tab. 10). The lengths of exons and introns are shown in Tab. 10. The longest exon is exon 1, containing 885 nucleotides, and the shortest exons are exon 10 and 16, containing 51 nucleotides each.

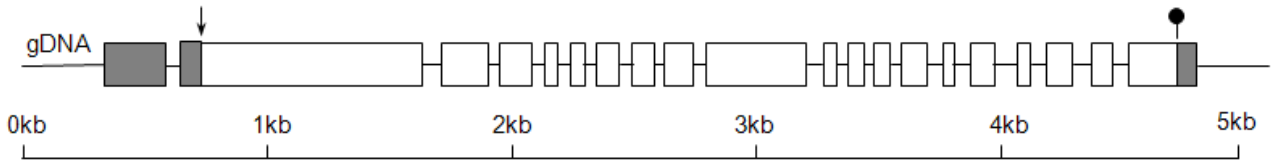


Fig. 22 Molecular map of the genomic DNA of *SmRpoT*. Nineteen exons are shown as boxes indicating the coding region. The translation initiation site (AUG) is designated by an arrow and the stop codon by a black circle. The 5' and 3' untranslated regions are designated by hatched areas. The grey lines between the exons indicate the 18 introns.

Tab. 10 Length of the introns and the exons of *SmRpoT* gene

Intron	length (bp)		Exon	length (bp)
5'-UTR	373			
Intron 0	67		Exon 1	885
Intron 1	54		Exon 2	201
Intron 2	60		Exon 3	129
Intron 3	64		Exon 4	63
Intron 4	53		Exon 5	59
Intron 5	53		Exon 6	118
Intron 6	48		Exon 7	114
Intron 7	54		Exon 8	132
Intron 8	63		Exon 9	414
Intron 9	51		Exon 10	51
Intron 10	55		Exon 11	71
Intron 11	59		Exon 12	73
Intron 12	53		Exon 13	153
Intron 13	63		Exon 14	75
Intron 14	51		Exon 15	90
Intron 15	69		Exon 16	51
Intron 16	64		Exon 17	115
Intron 17	53		Exon 18	104
Intron 18	65		Exon 19	108
SmRpoT: 4551 bp			3'-UTR	73

3.1.2 Molecular cloning of the RpoT genes and cDNAs of *Nuphar advena*

The identification and isolation of the *RpoT* genes from *Nuphar advena* followed the same scheme as described for *Selaginella*. In brief, nine commercially available filters of a *Nuphar advena* BAC library were subjected to hybridization with the same 1.5 kb probe used for *Selaginella*. Positive clones were identified and purchased for subsequent DNA sequence analysis to obtain the genomic DNA and, using RACE, the cDNA sequences.

3.1.2.1 Southern hybridization of the *N. advena* BAC filters

Hybridization of the *Nuphar advena* BAC library (nine filters) with the heterologous *RpoT*-cDNA probe from *Selaginella* (see "Materials and Methods") under non-stringent conditions resulted in the identification of 29 positive clones (Na-A to Na-I, see Fig. 23), with the number, name and coincident identified positions in the BAC library listed in Tab. 11. The 29 positive BAC clones were purchased for BAC DNA isolation and further analysis. Fig 24 shows twenty-four isolated BAC DNAs. For five of the 29 BAC clones (Nr. 19, 22, 24, 28 and 29) no DNA could be isolated, probably due to growth problems of the colonies.

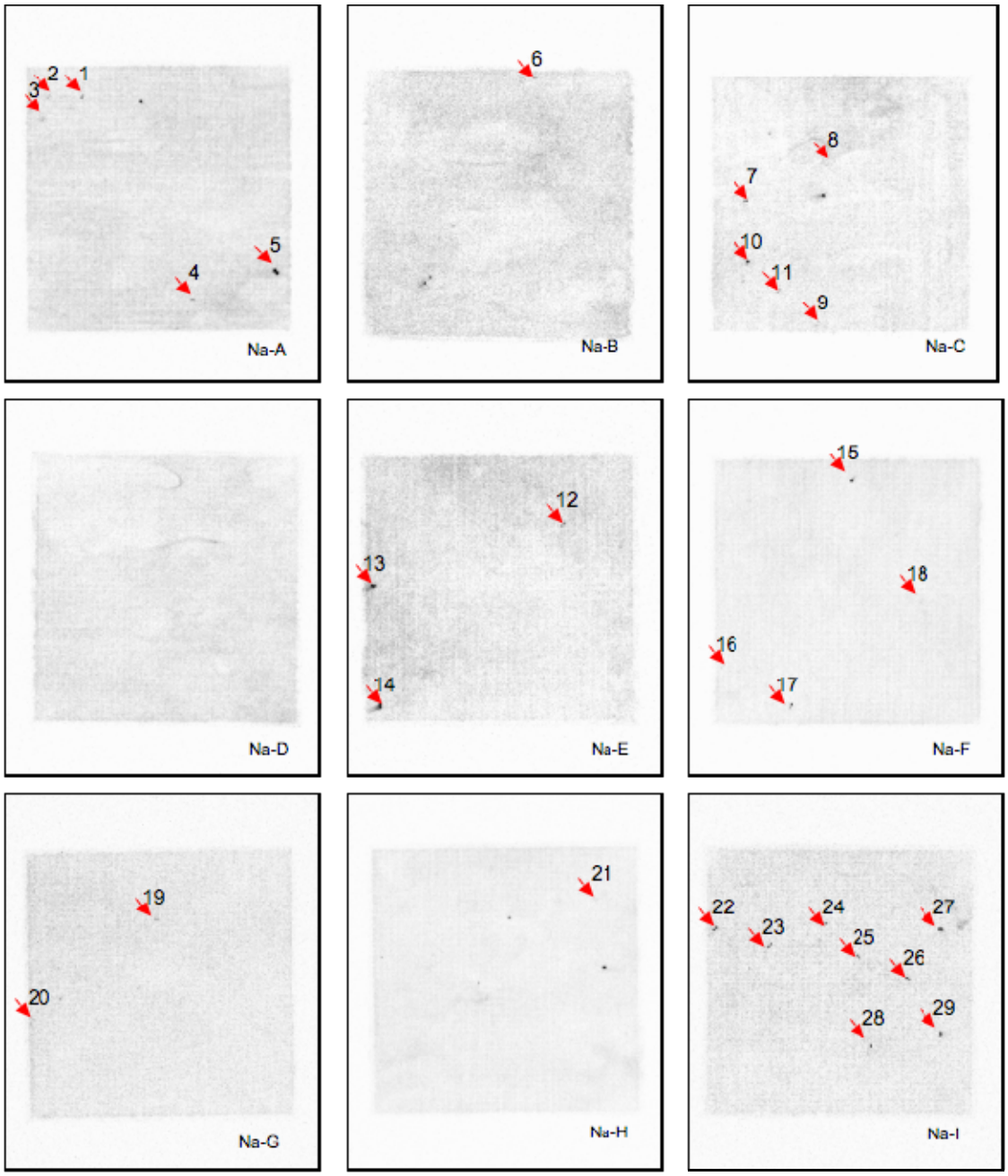


Fig. 23 Twenty-nine positive clones were identified from nine *Nuphar* BAC filters. The 29 positive clones, identified by BAC filter hybridization, are indicated by red arrows and their names as well as their positions on the BAC filters (see Tab. 11).

Tab. 11 Twenty-nine positive clones were identified from *Nuphar* BAC filters.

Clone Nr.	Clone name	Position
1	Na_BAC_A1	NA—A—Field2—F6—--pos5—plate 26
2	Na_BAC_A2	NA—A—Field2—L6—--pos5—plate 26
3	Na_BAC_A3	NA—A—Field2—M10—pos5—plate 26
4	Na_BAC_A4	NA—A—Field1—B19—pos4—plate 19
5	Na_BAC_A5	NA—A—Field5—C14—pos1—plate 5
6	Na_BAC_A6	NA—B—Field6—B2—-pos4—plate 72
7	Na_BAC_A7	NA—C—Field2—J24—-pos4—plate 116
8	Na_BAC_A8	NA—C—Field6—J17—-pos6—plate 132
9	Na_BAC_A9	NA—C—Field1—M23—pos6—plate 127
10	Na_BAC_A10	NA—C—Field4—J11—-pos8—plate 142
11	Na_BAC_A11	NA—C—Field4—D17—pos3—plate 112
12	Na_BAC_A12	NA—E—Field3—M13—pos3—plate 207
13	Na_BAC_B1	NA—E—Field2—O24—pos4—plate 212
14	Na_BAC_B2	NA—E—Field4—N22—pos2—plate 202
15	Na_BAC_B3	NA—F—Field6—H4—-pos8—plate 288
16	Na_BAC_B4	NA—F—Field4—O14—pos5—plate 268
17	Na_BAC_B5	NA—F—Field4—C21—pos7—plate 280
18	Na_BAC_B6	NA—F—Field5—K2—--pos3—plate 257
19	Na_BAC_B7	NA—G—Field6—I13—-pos4—plate 312
20	Na_BAC_B8	NA—G—Field4—P7—--pos6—plate 322
21	Na_BAC_B9	NA—H—Field3—H10—pos4—plate 357
22	Na_BAC_B10	NA—I—-Field2—O15—pos3—plate 398
23	Na_BAC_B11	NA—I—-Field2—E18—pos3—plate 398
24	Na_BAC_B12	NA—I—-Field6—K14—pos4—plate 408
25	Na_BAC_C1	NA—I—-Field6—E20—pos1—plate 390
26	Na_BAC_C2	NA—I—-Field3—L24—pos1—plate 387
27	Na_BAC_C3	NA—I—-Field3—F15—pos4—plate 405
28	Na_BAC_C4	NA—I—-Field1—C12—pos7—plate 421
29	Na_BAC_C5	NA—I—-Field5—F10—pos2—plate 395

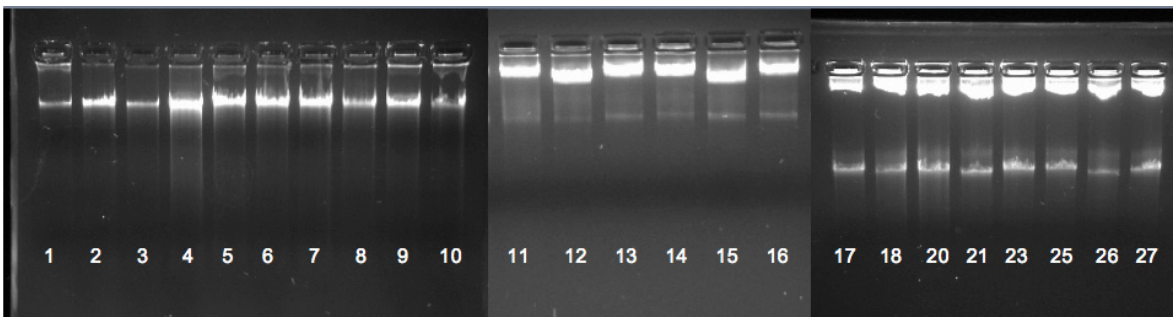


Fig. 24 Twenty-four *Nuphar advena* BAC DNAs were isolated and purified. The indicated numbers were coincident with the clone numbers in Tab. 11.

3.1.2.2 Identification of the genomic DNA sequences of three *NaRpoT* genes

Trace sequence data of *Nuphar advena* from NCBI (http://www.ncbi.nlm.nih.gov/Traces/) were extracted and aligned with the *RpoT* gene sequences from other plant species including *Selaginella moellendorffii* and *Arabidopsis taliana*. Two primers Na_seq_Tm_1 (CAG AGG TCT ATC CGC TGA AT) and Na_seq_Tm_2 (CTA GTG GGC GTC CTT CTG) (See Tab. 6) were designed as initial sequencing primers for obtaining first sequence tags of all 24 BAC DNAs.

The alignment of resulting sequences indicated three similar, but different sequences with high homology to *RpoT* genes, reflecting that there were at least three different *RpoTs* in *Nuphar advena*. Based on their homology with the *RpoT* genes from other plant species, these three *RpoTs* were named *NaRpoTm1*, *NaRpoTm2* and *NaRpoTp*. Primers were designed for these three *RpoT*s for sequencing both PCR sub-fragments and directly BAC DNA (for primers, see Tab. 6).

For *NaRpoTm1*, eight BAC clones, Na_BAC_A1, Na_BAC_A5 Na_BAC_A8, Na_BAC_A12, Na_BAC_B2, Na_BAC_B5, Na_BAC_B6 and Na_BAC_C1 (listed in Tab. 11), were subjected to BAC DNA primer walking and PCR with genomic DNA by using the corresponding primer pairs listed in Tab. 6. The *NaRpoTm1* genomic DNA sequence is 28,454 bp in length. Thus, the length of *NaRpoTm1* (as well as of the other *Nuphar RpoT* genes) is much longer than the *RpoT* gene from *Selaginella* (4.5 kb). Due to a few sequence gaps, the exact length of the *NaRpoTm2* gene is not known yet. However, it is at least 16.2 kb long. For *NaRpoTm2*, BAC clones Na_BAC_A2, Na_BAC_A11, Na_BAC_B3, Na_BAC_B4, Na_BAC_B8, Na_BAC_B9 and Na_BAC_B11 contributed to the acquisition of the gene sequence, while BAC clones Na_BAC_A3, Na_BAC_A4, Na_BAC_A6, Na_BAC_A9, Na_BAC_B1, and Na_BAC_C2 were used for sequencing the *NaRpoTp* gene, which is 13,555 bp in length. The sequences of the three *NaRpoT* genes were deposited in the EMBL database and are available under the accession numbers FN811768 (*NaRpoTm1*), FN820498 (*NaRpoTm2*) and FN811769 (*NaRpoTp*), respectively.

3.1.2.3 Identification of the full length *NaRpoT* cDNA sequences

Based on the obtained BAC sequences, 3'-RACE primers NaA-fw12, NaA-fw14, NaA-fw15 (for *NaRpoTm1*); NaC-fw4, NaC-fw5, NaC-fw6, NaC-fw7 (for *NaRpoTm2*); NaP-3R1, NaP-3R2, NaP-3R3 (for *NaRpoTp*) and 5'-RACE primers NaA-rev11, NaA-rev12, NaA-rev13, NaA-rev14 (for *NaRpoTm1*); NaC-rev16, NaC-rev17, NaC-rev18 (for *NaRpoTm2*); NaP-5R2, NaP-5R4, NaP-5R5, NaP-5R6, NaP-5R8 (for *NaRpoTp*) (see Tab. 6) were designed and applied to amplify the 3'- and 5' cDNA sequences of the three *NaRpoT* genes by PCR. Fig. 25 shows the RACE PCR results for *NaRpoTm1*. As expected, with every next nested PCR, the fragment sizes decreased according to the position of the primer targeting the cDNA sequences.

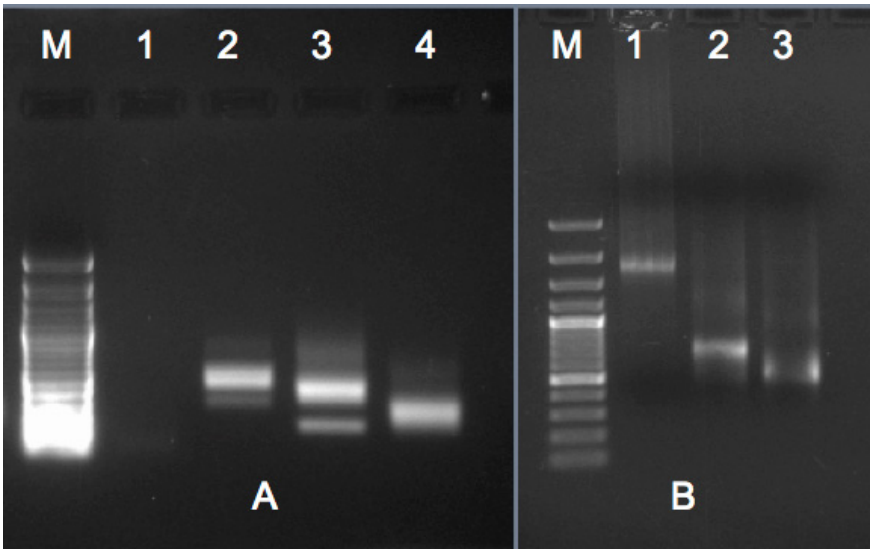


Fig. 25 RACE PCR of *NaRpoTm1* cDNA. A. 5'-RACE from four rounds of PCR (Lanes 1-4) B. 3'-RACE from three rounds of PCR (Lane 1-3). M, marker, 100 bp ladder plus.

After obtaining the sequences of the RACE fragments, full-length cDNA sequences were determined by performing overlapping amplifications with internal gene-specific primers (cDNA primers listed in Tab. 6). All fragments were sequenced and the subsequent contigs were assembled (SeqMan 7.2.1, DNASTAR). The lengths of cDNA sequences, protein coding regions, 5' untranslated leaders and 3' untranslated sequence of the three *NaRpoT* genes (*NaRpoTm1*, *NaRpoTm2* and *NaRpoTp*) are summarized in Tab. 12. In the case of *NaRpoTp*, no canonical translation start motif (ATG) was found at the 5' end of the cDNA. Instead, a CUG codon was found at position +148, from which translation could be initiated. Stop codons in the 5' region of this position excluded further upstream translation. Thus, it was assumed that translation of *NaRpoTp* starts from this CUG codon (see also 3.4.2.3.2.).

Tab. 12 Characteristics of full-length cDNAs of *NaRpoTs*

Gene	Full-length cDNA	5' untranslated leader	3' untranslated sequence	Coding sequence	Encoded amino acids
NaRpoTm1	3,671 nt	147 nt	536 nt	2,988 nt	996
NaRpoTm2	3,631 nt	294 nt	366 nt	2,970 nt	990
NaRpoTp	3,313 nt	93 nt	265 nt	2,955 nt	985

The structures of the three *NaRpoT* genes are shown in Fig. 26. Each of the three *NaRpoT* genes consists of 19 exons and 18 introns. The lengths of exons and introns of the three *NaRpoT* genes are shown in Tab. 13. The positions of all introns and exons of the three *Nuphar RpoT* genes are precisely conserved with those from other plant species (see Fig. 27), and as well, all exon-intron junctions contain conserved GT and AG sequences at the 5'- and 3'- ends of the introns, respectively.

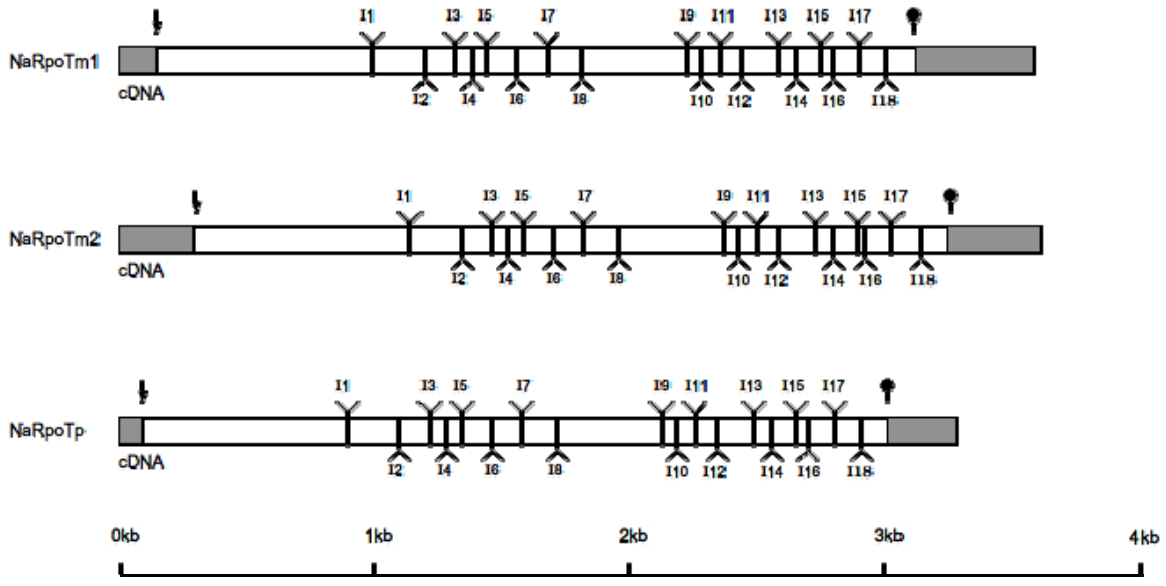


Fig. 26 Gene structures of three *NaRpoT* genes. Exons are shown as white boxes indicating the coding region, with the translation initiation sites (AUG for *NaRpoTm1* and *NaRpoTm2*, and a putative CUG start codon for *NaRpoTp*, see text) designated by an arrow and the stop codon by a black circle. The 5' and 3' untranslated regions are designated by hatched areas. Introns are numbered I1 through I18.

Tab. 13 Lengths of the introns and the exons of the three *NaRpoT* genes

NaRpoTm1 (28,454 bp)				NaRpoTm2 (>16,207 bp)				NaRpoTp (13,555 bp)			
Exon	Length (bp)	Intron	Length (bp)	Exon	Length (bp)	Intron	Length (bp)	Exon	Length (bp)	Intron	Length (bp)
5'-UTR	147			5'-UTR	294			5'-UTR	93		
Exon1	864	Intron1	183	Exon1	840	Intron1	176	Exon1	807	Intron1	1349
Exon2	210	Intron2	76	Exon2	210	Intron2	76	Exon2	207	Intron2	165
Exon3	128	Intron3	95	Exon3	134	Intron3	88	Exon3	134	Intron3	390
Exon4	67	Intron4	707	Exon4	67	Intron4	728	Exon4	67	Intron4	401
Exon5	59	Intron5	4338	Exon5	59	Intron5	384	Exon5	59	Intron5	756
Exon6	118	Intron6	6775	Exon6	118	Intron6	>2329	Exon6	118	Intron6	2086
Exon7	114	Intron7	372	Exon7	114	Intron7	209	Exon7	114	Intron7	134
Exon8	132	Intron8	84	Exon8	132	Intron8	85	Exon8	132	Intron8	105
Exon9	417	Intron9	1549	Exon9	417	Intron9	>1652	Exon9	417	Intron9	673
Exon10	51	Intron10	3883	Exon10	51	Intron10	>1648	Exon10	51	Intron10	1021
Exon11	71	Intron11	2805	Exon11	70	Intron11	>1432	Exon11	71	Intron11	1092
Exon12	85	Intron12	963	Exon12	86	Intron12	422	Exon12	85	Intron12	302
Exon13	150	Intron13	170	Exon13	149	Intron13	187	Exon13	153	Intron13	80
Exon14	75	Intron14	1864	Exon14	76	Intron14	2131	Exon14	75	Intron14	99
Exon15	90	Intron15	278	Exon15	90	Intron15	182	Exon15	90	Intron15	822
Exon16	48	Intron16	110	Exon16	30	Intron16	137	Exon16	48	Intron16	300
Exon17	115	Intron17	74	Exon17	115	Intron17	72	Exon17	115	Intron17	309
Exon18	104	Intron18	457	Exon18	104	Intron18	639	Exon18	104	Intron18	158
Exon19	108	3'-UTR	536	Exon19	108	3'-UTR	366	Exon19	108	3'-UTR	265

The *NaRpoTm1* genomic DNA sequence is 28,454 bp in length with some very long introns, e.g., the lengths of introns 5, 6 and 10 are 4,338 bp, 6,775 bp and 3,883 bp, respectively. Only 16.2 kb of the genomic sequence of *NaRpoTm2* has been obtained. This is not the complete genomic sequence of *RpoTm2*, because some very long intron sequences contain gaps and have not been completely finished (introns 6, 9, 10 and 11, as listed in Tab 13). Since the missing sequences are all exclusively located within introns, the position of the 19 exons and 18 introns with conserved exon-intron junctions could be unequivocally determined. The *NaRpoTp* gene is 13,555 bp in length. The longest exon is exon 1 with 807 nucleotides and the shortest one is exon 16 with 48 nucleotides. Compared with the other two *RpoT* genes (the length of each intron from *NaRpoTm1* and *NaRpoTm2* is 74-6,775 and 72- >4,733 nt, respectively), the lengths of the 18 introns (80-2,086 nt for each) of *NaRpoTp* are in general shorter. However, all the three *Na-RpoT* genes have similar sizes for each corresponding exon. Eleven of their exons have even identical sizes (exon 4 with 67 nt, exon 5 with 59 nt, exon 6 with 118 nt, exon 7 with 114 nt, exon 8 with 132 nt, exon 9 with 417

nt, exon 10 with 51 nt, exon 15 with 90 nt, exon 17 with 115 nt, exon 18 with 104 nt and exon 19 with 108 nt). The insertion sites of all introns are precisely conserved as well.

3.2 Sequence comparison of the RpoT polymerases from *Selaginella* and *Nuphar* with other plant RpoT polymerases

The protein coding sequence of *SmRpoT* is 3,006 bp long and the predicted SmRpoT protein comprises of 1,002 amino acids exhibiting an identity of 48.6–49.5% and 48.0–52.3%, with the *Arabidopsis* and *Physcomitrella patens* RpoT polypeptides, respectively. The protein coding sequences of *NaRpoTm1* and *NaRpoTm2* are 2,988 bp and 2,970 bp long, respectively, and the deduced proteins comprise of 996 and 990 amino acids, respectively. The third RpoT polymerase from *Nuphar*, NaRpoTp, comprises of 985 amino acids, with 63.1% and 64.6% identity to NaRpoTm1 and NaRpoTm2, respectively. The alignment of the RpoT polymerases from *Selaginella* and *Nuphar* with those from *Arabidopsis* and *Physcomitrella* (Fig. 27) demonstrates a high degree of conservation at the amino acid sequence level (Tab. 14), most striking in the C-terminal part, and including all functionally crucial regions and residues known from the phage T7 RNA polymerase (Mcallister & Raskin, 1993; Sousa *et al.*, 1993).

Tab. 14 Similarity comparison of RpoT polymerases between *Arabidopsis thaliana* (At), *Physcomitrella patens* (Pp), *Selaginella moellendorffii*(Sm) and *Nuphar advena* (Na).

	Percentage of identity												
		1	2	3	4	5	6	7	8	9	10		
	1	-----	61.6	55.1	50.1	49.6	46.2	49.5	62.4	63.8	60.2	1	AtRpoTm
	2	53.4	-----	55.2	50.3	52.0	47.4	48.6	64.7	65.6	61.2	2	AtRpoTmp
	3	67.1	66.9	-----	50.1	51.2	46.9	48.8	57.8	58.7	60.4	3	AtRpoTp
	4	79.6	79.1	79.7	-----	59.3	56.2	52.3	52.0	52.7	51.8	4	PpRpoT1mp
	5	80.8	74.6	76.8	58.0	-----	53.8	51.7	52.5	53.2	52.3	5	PpRpoT2mp
	6	90.5	87.1	88.5	64.6	70.2	-----	48.0	46.5	47.4	47.3	6	PpRpoT3
	7	81.1	83.6	83.1	74.0	75.5	85.4	-----	50.5	51.1	49.5	7	SmRpoTm
	8	51.8	47.4	61.3	74.6	73.4	89.7	78.6	-----	96.8	63.1	8	NaRpoTm1
	9	49.1	45.8	59.3	73.0	71.8	87.1	76.9	3.3	-----	64.6	9	NaRpoTm2
	10	56.2	54.1	55.7	75.1	73.9	87.3	81.7	50.5	47.7	-----	10	NaRpoTp
Divergence		1	2	3	4	5	6	7	8	9	10		

Regarding the gene structure, the insertion sites of all common introns are precisely conserved relative to the aligned amino acid sequences between the *SmRpoT* and *NaRpoT* genes and those of *Arabidopsis* and *Physcomitrella* (Fig. 27). All exon-intron junctions contain conserved GT and AG sequences at the 5'- and 3'- ends of the introns, respectively. Compared with the *RpoT* genes from *Arabidopsis*, *Physcomitrella* and *Nuphar*, *SmRpoT* contain one additional intron (intron 0) in the 5' part of the gene upstream of the putative initiation code. Both *Selaginella* and *Nuphar* genes miss

the additional introns (one of *PpRpoT1* and two of *PpRpoT2*) which exist in the 5' part of the *Physcomitrella RpoT* genes. In the *Selaginella* gene, the lengths of the 18 introns (48-69 nt) are in general shorter than those of *Arabidopsis* (52-612 nt), (Hedtke *et al.*, 1997; Hedtke *et al.*, 2000) and *Physcomitrella* (120-266 nt), (Richter *et al.*, 2002), in agreement with the fact that *Selaginella moellendorffii* has the smallest genome (only ~100 Mbp) among all plant species investigated thus far (Wang *et al.*, 2005; Weng *et al.*, 2005), (http://genome.jgi-psf.org/Selmo1/Selmo1.home.html). The *Nuphar RpoT* genes contain a number of very long introns (for sizes of introns see Tab 13), corresponding with the much larger genome size (3,040 Mbp) of this species (http://www.genome.arizona.edu/BAC_ special_projects/). With 985 amino acids, NaRpoTp shows 63.1% and 64.6% identity to both NaRpoTm1 and NaRpoTm2.

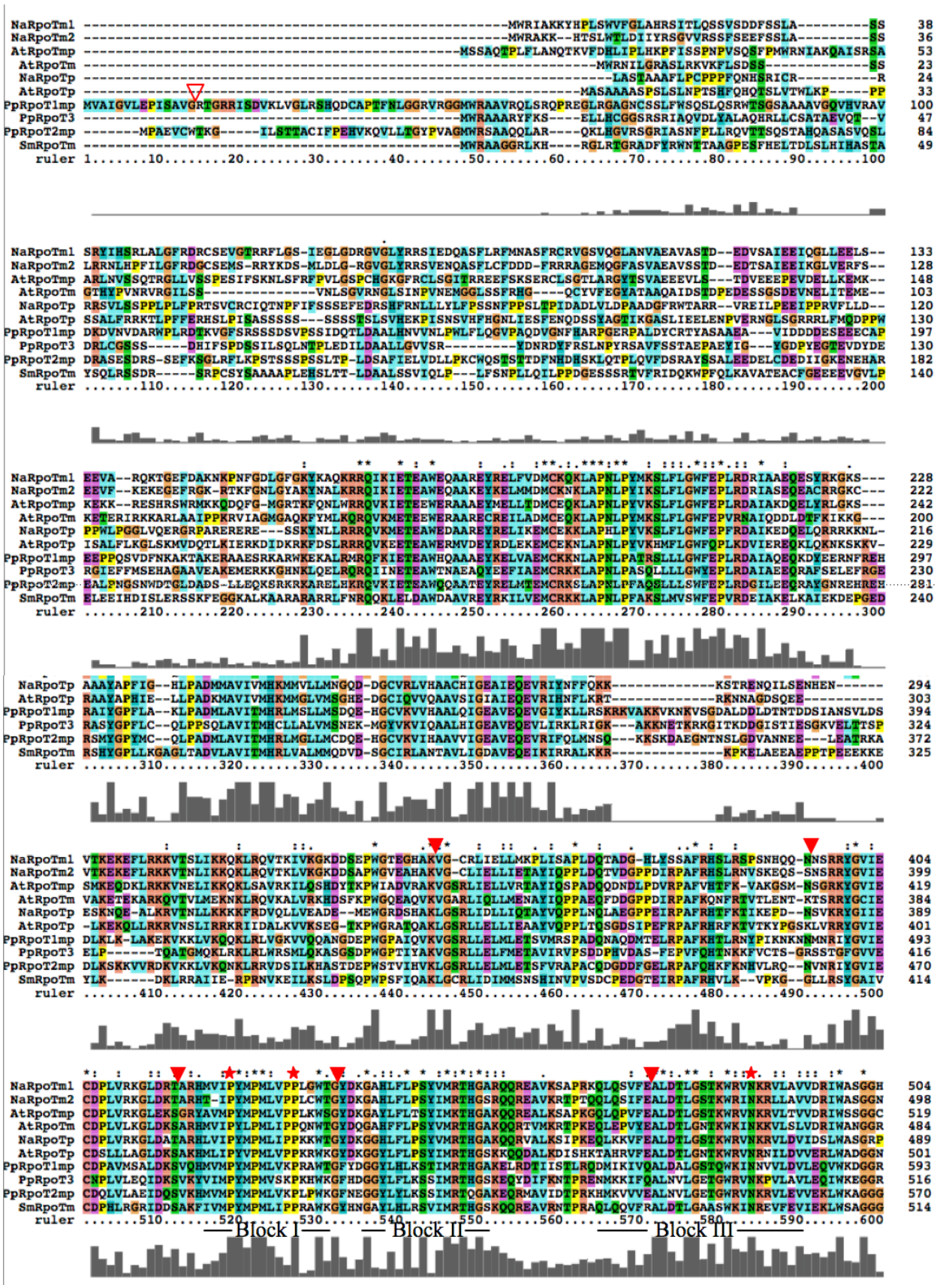
3.3 *RpoT* gene copy number

From representativ BAC libraries, one (*Selaginella*) and three (*Nuphar*) *RpoT* genes were isolated, respectively. Experiments were performed to investigate whether the number of isolated *RpoT* genes corresponds to the real gene copy number. For this purpose, Southern hybridization of total genomic DNA from both organisms with a partial *RpoT* cDNA sequence (for *Selaginella moellendorffii*) or a set of three partial *RpoT* cDNA sequences (for *Nuphar advena)* as probes was performed under non-stringent conditions to detect putative additional copies of *RpoT* genes. The probes were derived from the 3'-*RpoT* region which is highly conserved between *RpoT* sequences across species.

3.3.1 RpoT gene copy number in *Selaginella moellendorffii*

3.3.1.1 Selection of probe and restriction cleavage sites

A 1,518 bp cDNA fragment amplified by primer pair 'Smin-p' and 'Smin-m1' was used as hybridization probe. The target sequence is located in the 3' conserved part of the whole gene, located between exon 6 and 3' UTR. This region is 2.2 kb in length, including 14 exons, 13 introns and the 3' UTR. Several enzyme cleavage sites were found in the flanking regions of the probe's target sequence, upstream and downstream (see Fig. 28). Restriction sites for BamHI, DraI, and BstXI were selected and expected to produce cleavage products of 2,800 bp, 4,632 bp and 4,558 bp, respectively.


Block I
Block II
Block III

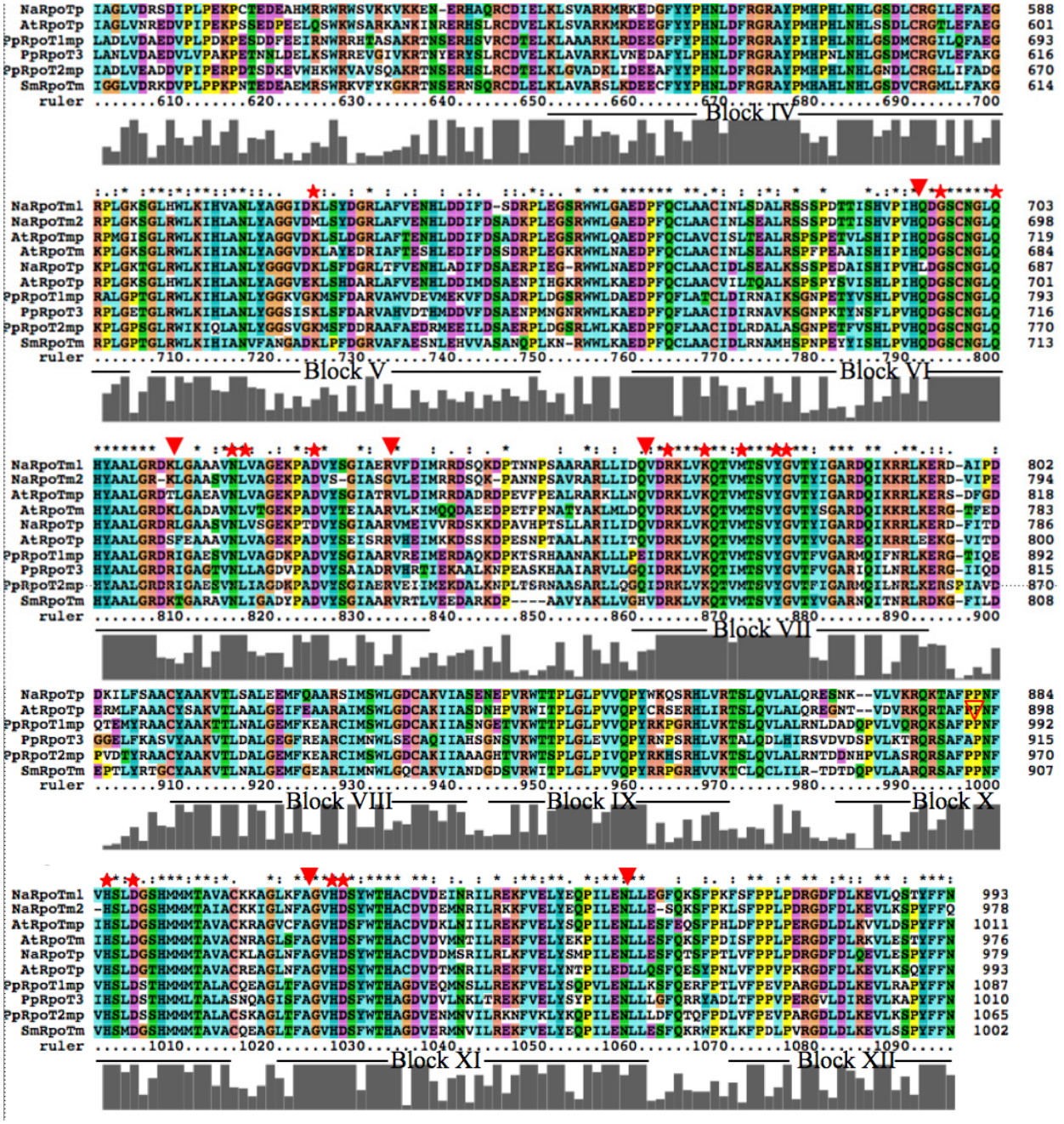


Fig. 27 Comparison of the deduced amino acid sequences of 10 RpoT polymerases. The amino acid sequences of the RpoTs from *Nuphar* (NaRpoTm1, NaRpoTm2 and NaRpoTP), *Selaginella* (SmRpoTm), *Arabidopsis* (AtRpoTm, AtRpoTp and AtRpoTmp) and *Physcomitrella* (PpRpoT1mp, PpRpoT2mp and PpRpoT3) were aligned using ClustalW. Accession numbers are as follows: AtRpoTm, P92969; AtRpoTmp, CAC17120; AtRpoTp, O24600; PpRpoTmp1, CAC95163; and PpRpoTmp2, CAC95164. PpRpoT3 is an RpoT amino acid sequence derived from the database of the *Physcomitrella patens* genome project (http://genome.jgi-psf.org/phypa1_1/phypa1_1. home.html, scaffold_241:459109-464235). In silico analysis of the genome as well as expressed sequence tag (EST) data strongly suggest that the sequence, designated as PpRpoT3, is a product of an *RpoT* gene with the conserved intron–exon structure of land plants that encodes a functional RNA polymerase (U. Richter, unpublished data). Black lines indicate conserved blocks in the RpoT polymerase family; functionally crucial residues (Mcallister & Raskin, 1993; Sousa *et al*., 1993) are indicated by asterisks. The position of common introns is designated by filled triangles and PpRpoT2mp-specific introns by open triangles. Conserved amino acid positions (60%) are shaded.

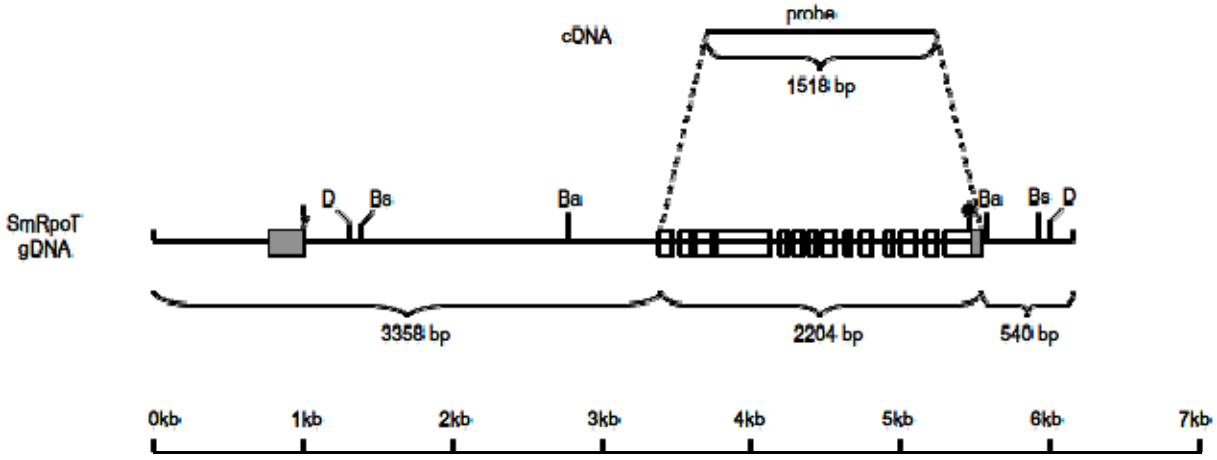


Fig. 28 Molecular map of *SmRpoT* and probe position. Exons are shown as boxes indicating the coding region, with the translation initiation site (AUG) designated by a horizontal arrow and the stop codon by a black circle. The 5'- and 3'- untranslated regions are designated by hatched areas (grey bar). Restriction sites for BamHI (B), DraI (D), and BstXI (Bs) are indicated. The position of the 1.5-kb fragment used as a hybridization probe is indicated by a thin line.

3.3.1.2 Southern hybridization

Total DNA from *Selaginella* was digested with BamHI, DraI, and BstX1, respectively. The fragments were electrophoretically separated, blotted to a Nylon membrane and hybridized under non-stringent conditions with the 1.5 kb probe described above. Fig. 29 shows the results of the Southern hybridization. In each of the three restriction digests, one band was detected with the size corresponding to the predicted length of the RpoT fragment. Under non-stringent hybridization conditions, no additional bands were detected, indicating that *Selaginella* encodes only one *RpoT* gene. This observation is supported by data available from the *Selaginella* genome project (http://www.phytozome.net/selaginella.php). In the sequence database of the project, although not completely assembled, we identified only one single *RpoT* sequence which exactly coincides with the *RpoT* gene sequence reported in this study.

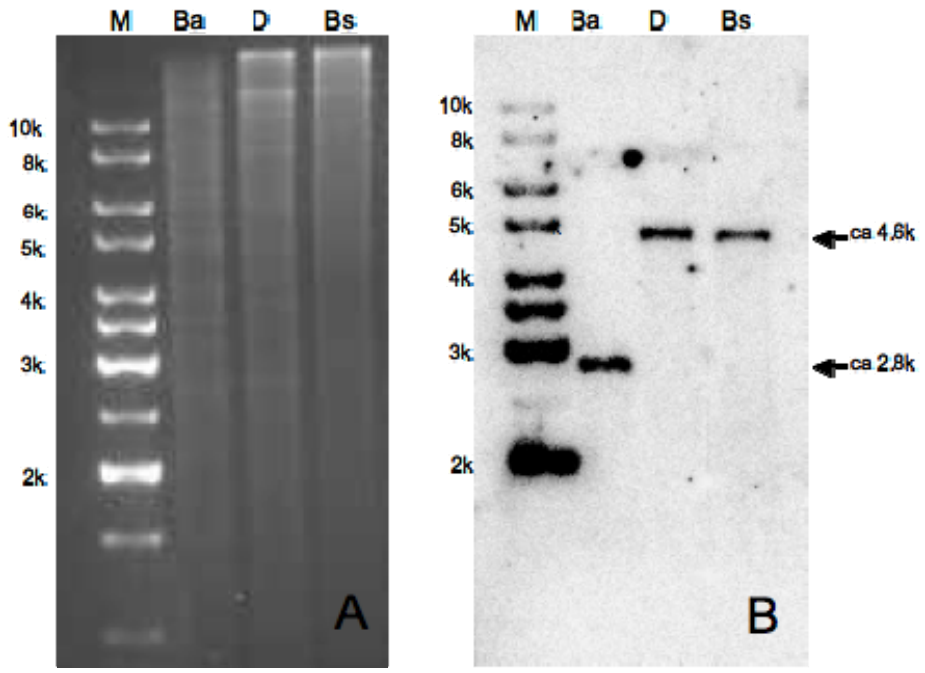


Fig. 29 Only one *RpoT* gene was identified in *Selaginella*. A. Genomic DNA of *Selaginella* was digested with restriction enzymes BamHI (Ba), DraI (D), and BstXI (Bs). B. The DNA was transferred to nylon membrane and hybridized with a 32p-labeled 1.5-kb cDNA probe (position of the probe shown in Fig. 28). M: size marker, 1 kb ladder.

3.3.2 *RpoT* gene copy number in *Nuphar advena*

3.3.2.1 Selection of probe

The Southern hybridization probe for *Nuphar advena* comprised of three individual cDNA fragments from the three *NaRpoT* genes, in order to allow detection of putative additional *RpoT*-related sequences specifically for all of the three known genes. The lengths of the three probes were 571 bp for *NaRpoTm1*, 551 bp for *NaRpoTm2* and 597 bp for *NaRpoTp,* primer pairs used for amplification were 'NaA-sp2p'-'NaA-sp2m', 'NaC-sp1p'-'NaC-sp1m' and 'NaP-sp1p'-'NaP-sp1m', respectively (see Tab. 6). All 3 probe sequences are located in the 3' conserved parts of the genes, from exon 15 to exon 18, as shown in Fig. 30.

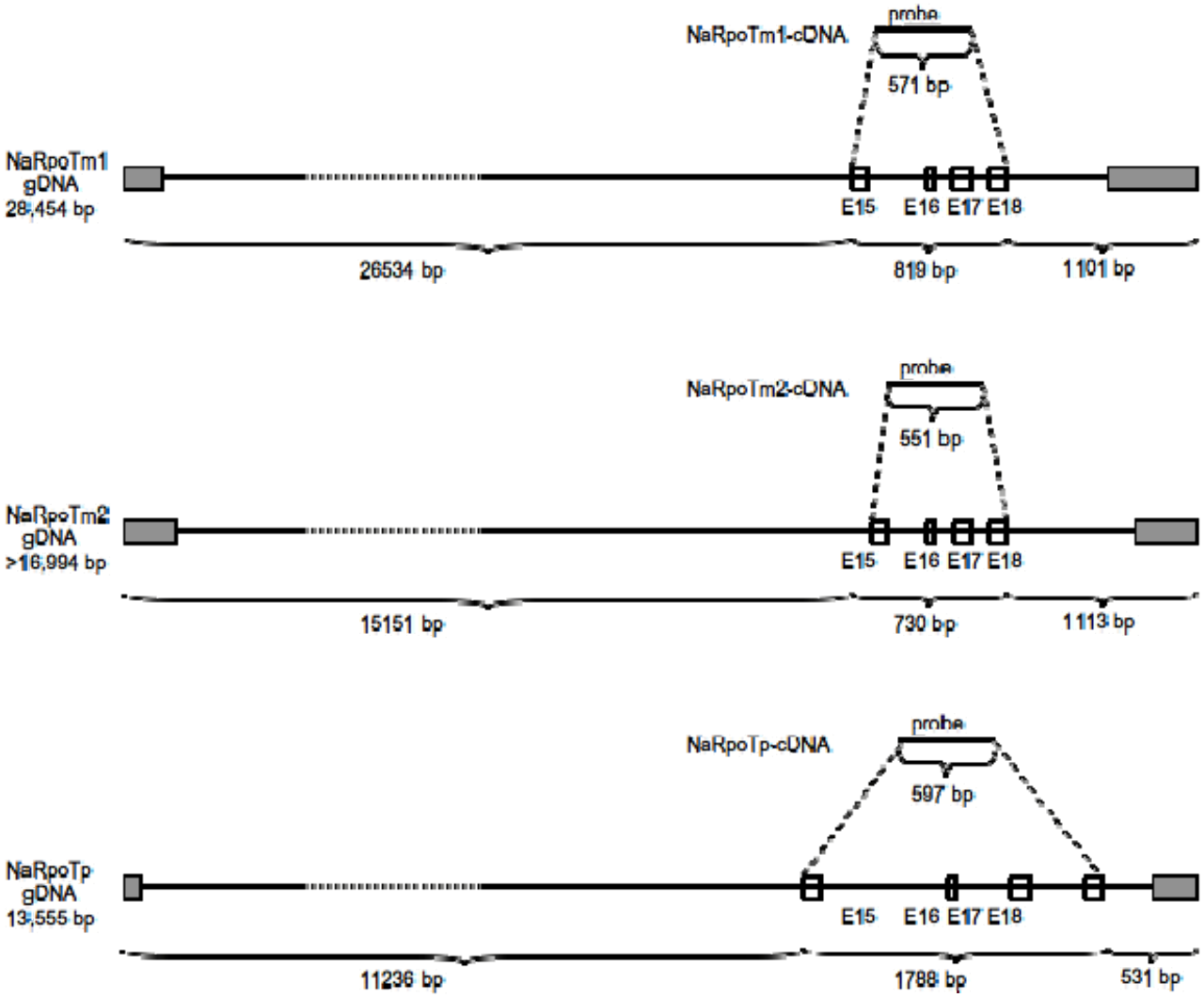


Fig. 30 Molecular map of three *NaRpoT* genes and the positions for probe. Exons 15-18 (E15-18) are shown as boxes, the 5' and 3' untranslated regions are designated by hatched areas (grey bar). The positions of the cDNA fragments used as hybridization probes are indicated by a thin line. For each *NaRpoT* gene (*NaRpoTm1*, *NaRpoTm2* and *NaRpoTp*), the lengths of the three regions (upstream of exon 15, from exon 15 to exon 18, downstream of exon 18) are indicated.

3.3.2.2 Restriction cleavage sites

Based on the genomic sequence, several restriction cleavage sites were found in or adjacent to the genomic DNA regions covered by the probes (Fig. 31). Restriction enzymes ScrFI, DraI, and BseLI were chosen to digest the genomic DNA. For each enzyme, the corresponding genomic DNA regions and lengths of the resulted product are shown for the three *NaRpoT* genes in Fig. 31.

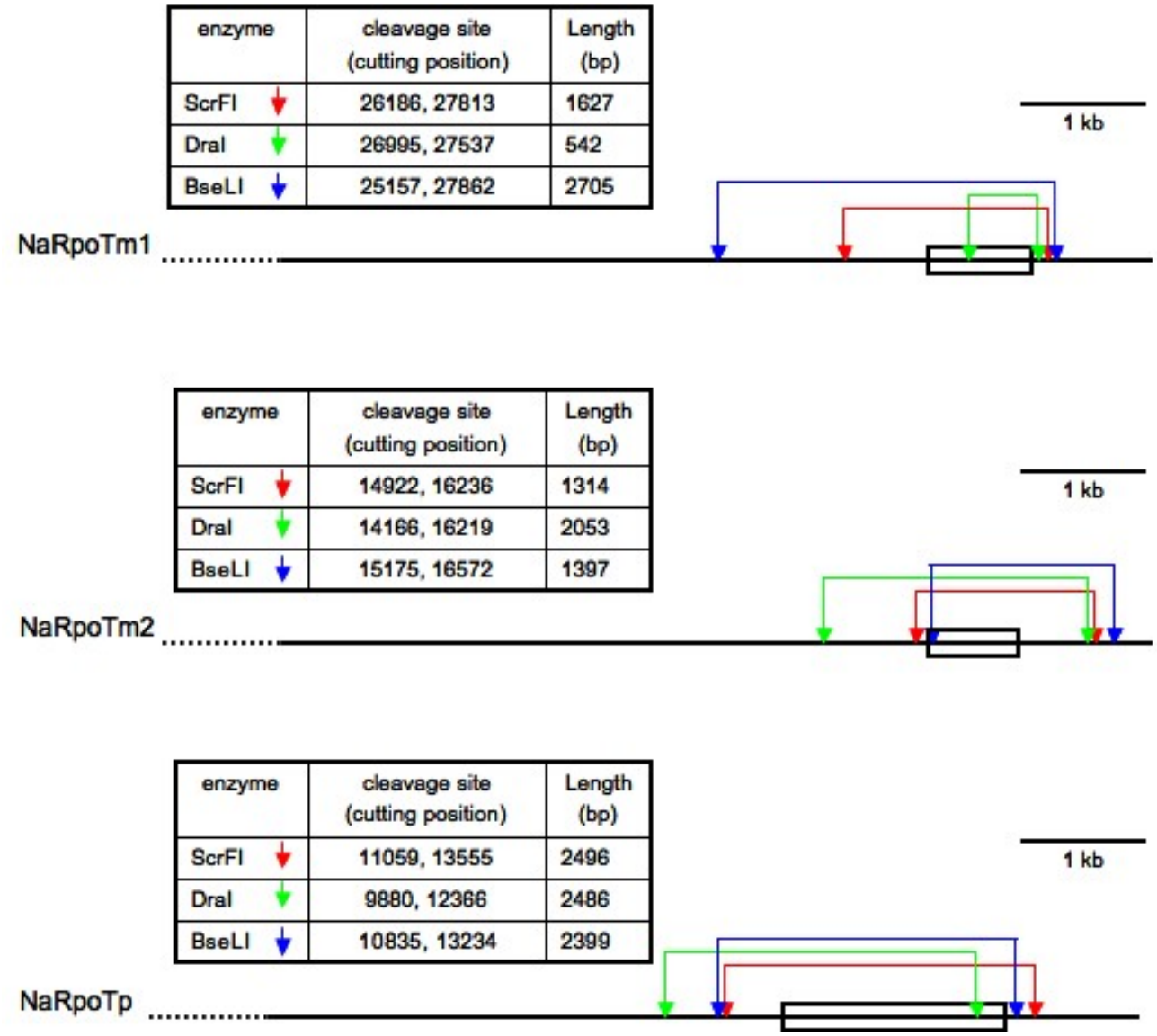


Fig. 31 *In silico* restriction analysis of the three *NaRpoT* genes. Unhatched boxes show the probe regions on the genomic DNA of the 3 *NaRpoT* genes. Red arrows indicate the cleavage sites of enzyme ScrFI; green arrows show the enzyme DraI and blue arrows the enzyme BaeLI. In all *NaRpoT* genes, cleavage with each enzyme (ScrFI, DraI and BaeLI) in or near the probe region produces only one specific restriction fragment, with the spanning regions and expected lengths listed in the inserted tables.

3.3.2.3 Southern hybridization

Total DNA from *Nuphar* was digested with ScrFI, DraI, and BseLI, respectively. The fragments were electrophoretically separated, blotted to a Nylon membrane and hybridized with the composite probe described above. Fig. 32 shows the results of the Southern hybridization. In each of the three restriction digests, three bands were detected with their sizes corresponding to the predicted lengths of the three *RpoT* gene fragments (Fig. 32). Additionally, weak and smeared hybridization signals were observed in all three restriction digests. Rather than indicating the presence of an additional *RpoT* gene, they most probably resulted from incomplete digestion of *Nuphar* DNA, whose enormous genome size of approximately 3,040 Mb made restriction analysis a hard work. From the

results of the Southern hybridization under non-stringent conditions we conclude that *Nuphar advena*, most likely, contains three *RpoT* genes.

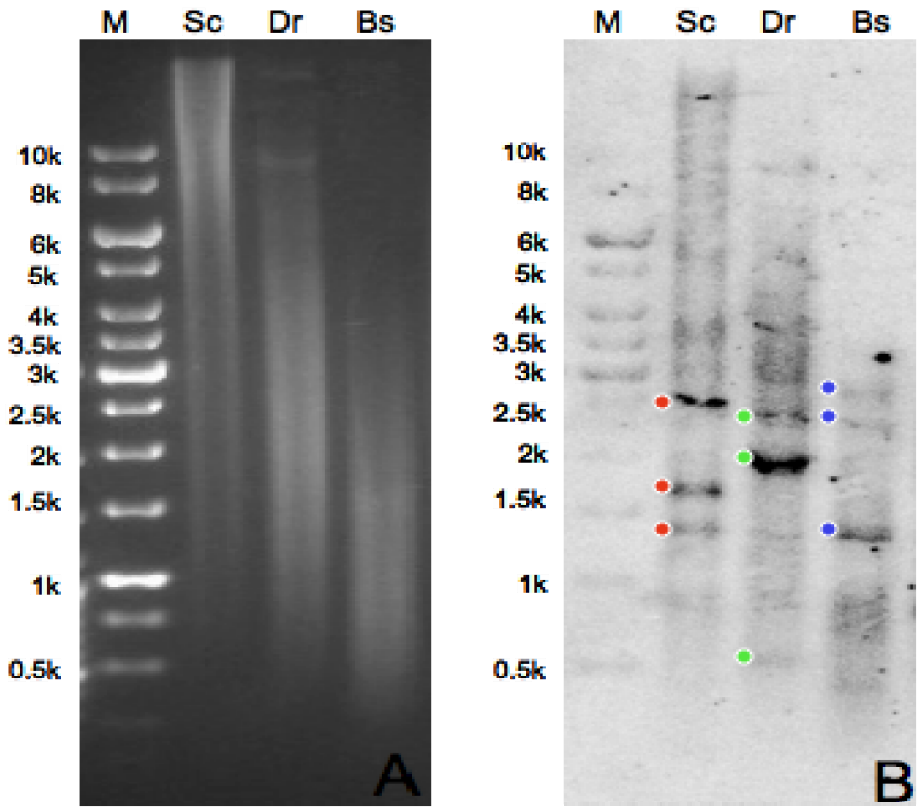


Fig. 32 Southern hybridization of *N. advena* genomic DNA with *RpoT*-specific probe. A. Genomic DNA of *Nuphar* was digested with restriction enzymes ScrFI (Sc), DraI (Dr), and BseLI (Bs). B. After gel electrophoresis, the DNA was transferred to nylon membrane and hybridized with a composite probe consisting of three 32p-labeled cDNA probes (positions of the probes shown in Fig. 31). Bands with predicted sizes marked by colored dots; M: marker, 1 kb ladder.

3.4 Subcellular localization of RpoT polymerases from *Selaginella* and *Nuphar*

3.4.1 *In silico* prediction of subcellular localization

Nuclear *RpoT* genes encode RNA polymerases, which are imported into organelles - mitochondria or plastids. The first information on the subcellular localization of the proteins, encoded by the *RpoT* genes identified in the present study, can be obtained by *in silico* analysis of their N termini (transit peptides) using appropriate prediction algorithms. Two programs were employed: TargetP 1.1 (Emanuelsson *et al.*, 2000) and Predotar (Small *et al.*, 2004) using neutral network algorithms. The amino termini of the deduced amino acid sequences of *SmRpoT* and the three *NaRpoTs* were analyzed by these programs, with the results of the prediction shown in Tab. 15. The RpoT polymerase from *Selaginella,* by both programs was predicted, with high probability (score

0.898/0.64), to be mitochondrially located. Two of the RpoTs from *Nuphar* were predicted also to be imported into mitochondria, one of them, however, with a relatively low probability score (0.588/0.28). For the third RpoT polymerase from *Nuphar*, both programs clearly specified plastid localization with a high score (0.988/0.73). Based on the *in silico* prediction results and experimental import studies (see below), the *RpoT*-encoded polymerases were named SmRpoTm, NaRpoTm1, NaRpoTm2 and NaRpoTp, respectively.

Tab. 15 *In silico* prediction of subcellular localization of the RpoT polymerases with TargetP and Predator. Mitochondrial: mitochondrial localization; Plastid: plastid localization; Elsewhere, neither in mitochondria nor in plastids. Values in brackets: probability scores.

	Target P			Predotar		
	Mitochondrial	Plastid	Elsewhere	Mitochondrial	Plastid	Elsewhere
SmRpoT	0.898	0.022	0.215	0.64	0.00	0.34
NaRpoTm1	0.608	0.063	0.041	0.71	0.00	0.26
NaRpoTm2	0.588	0.201	0.099	0.28	0.00	0.68
NaRpoTp	0.227	0.988	0.019	0.08	0.73	0.25

3.4.2 Experimental confirmation of subcellular localization of RpoTs

To verify the subcellular localization experimentally, the N termini of the RpoT protein sequences were translationally fused to GFP (Fig. 12 for SmRpoT, Fig. 14, Fig. 15 and Fig. 16 for NaRpoTs, as described in "Materials and Methods"). The fusion constructs were transiently expressed in protoplasts isolated from *Arabidopsis* and *Selaginella* for SmRpoT, and in *Arabidopsis* protoplasts for the three NaRpoTs. Subcellular localization of the GFP fusion constructs was observed using confocal laser scanning microscopy.

3.4.2.1 Control constructs for subcellular localization

Transformation with the mitochondrial coxIV-GFP (Akashi *et al.*, 1998) resulted in an accumulation of GFP in mitochondria, in punctuate structures of about 1 µm size (Fig. 33A, Fig. 33B and Fig. 24A) (Hedtke *et al.*, 1999; Hedtke *et al.*, 2000). A GFP fusion of the amino-terminus of *Arabidopsis* RecA (Akashi *et al.*, 1998) was used as a plastid control resulting in a coincidence of green GFP and red chlorophyll autofluorescence in the chloroplasts (see Fig. 33C, Fig. 33D and Fig. 34B). As expected, transformation of the protoplasts with the pOL plasmid containing no transit peptid resulted in cytosolic localization of GFP fluorescence (see Fig. 33E, Fig. 33F and Fig. 34C).

3.4.2.2 Subcellular localization of *Selaginella moellendorffii* RpoT

The subcellular localization of three different RpoT constructs (as outlined in Fig. 12) was examined: *SmRpoT$_{met}$-GFP* with the first encoded methionine cloned immediately downstream of the 35S promoter for forced translation, *SmRpoT$_{utr}$-GFP* containing the whole 5'untranslated region, and *SmRpoT$_{mut}$-GFP*, in which the methionine had been mutated to isoleucine (see also Fig. 13). To take into account the possible incompatibilities between transit peptide sequences from the spikemoss *Selaginella* and the import machineries of angiosperms, targeting of all constructs were examined not only in *Arabidopsis*, but also in *Selaginella* protoplasts. GFP fluorescence was observed by laser scanning microscopy. The transit peptides of the *Selaginella* RpoT as well as the control constructs successfully conferred import into organelles. In the case of *SmRpoT$_{met}$-GFP*, where translation is forced to start from the first encoded methionine, green GFP fluorescence was exclusively found in mitochondria (Fig. 33G and H). The same result was observed, when the construct contained the full length of the 5' untranslated leader sequence (*SmRpoT$_{utr}$-GFP, Fig. 33*I and J). About 10% of the transgenic protoplasts showed bright green fluorescence in these two constructs. In contrast, when the methionine site was substituted by isoleucine (construct *SmRpoT$_{mut}$-GFP*), importing of the GFP to mitochondria was abolished, and the cells did not show any green fluorescence (Fig. 33K and L). As can be seen from Fig. 33, the fluorescence characteristics of all constructs examined was the same in both *Arabidopsis* and *Selaginella* protoplasts. The experimental data indicated that the AUG at position +306 is the only available *RpoT* start codon, most likely used for translation initiation also *in planta* and specifying a transit peptide with strict mitochondrial targeting properties.

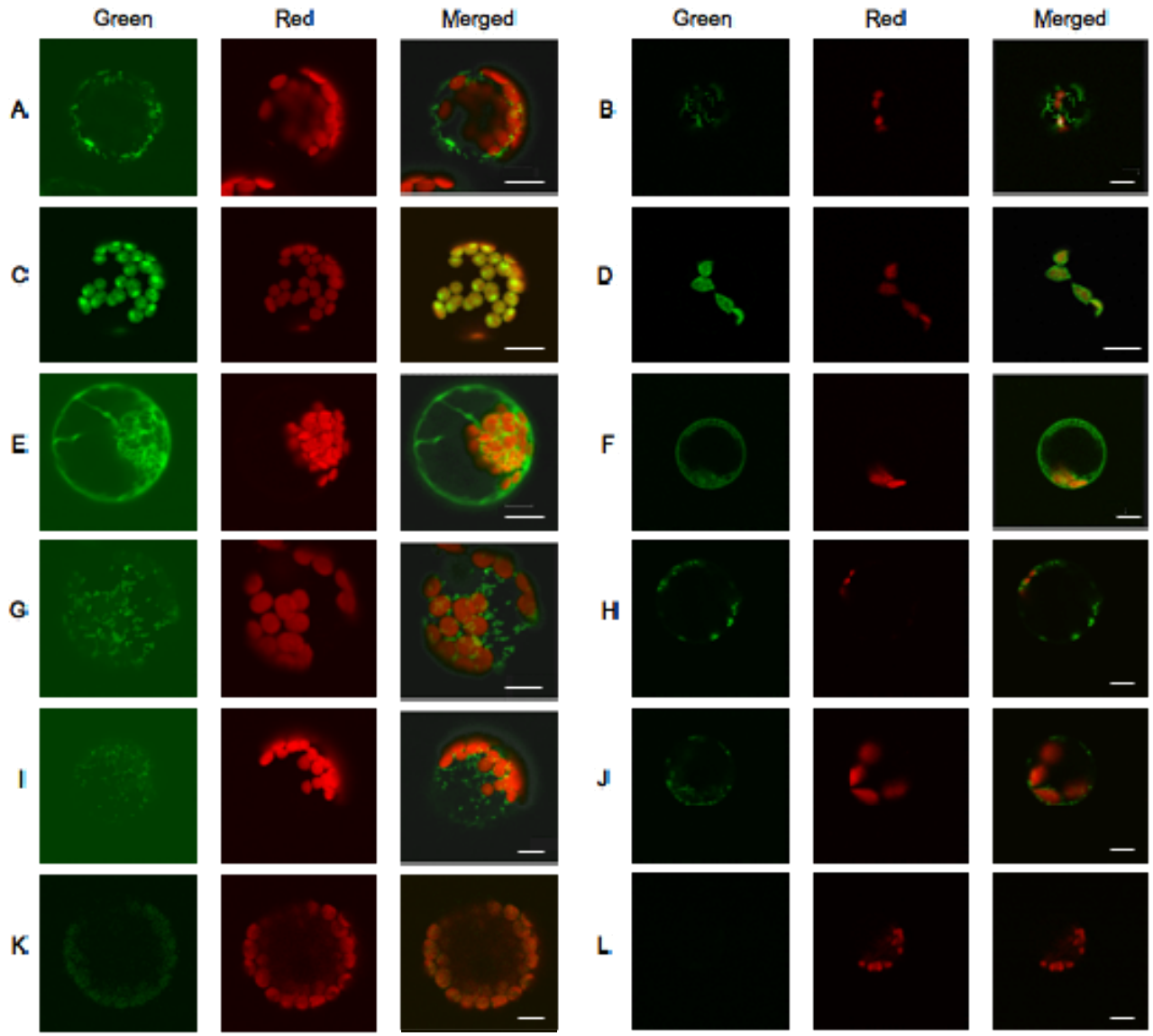


Fig. 33 Subcellular localization of SmRpoT by confocal laser scanning microscopy of transformed *Arabidopsis* (left panel) and *Selaginella* (right panel) protoplasts. The images depict fluorescence patterns (green, red and merged channels) of control constructs targeting GFP to mitochondria (A, B), plastids (C, D), vector control containing no transit peptide (E, F), SmRpoTmet-GFP (G and H), SmRpoTutr-GFP (I and J), and SmRpoTmut-GFP (K and L). Scale bar = 10 µm.

3.4.2.3 Subcellular localization of the three *Nuphar advena* RpoTs

3.4.2.3.1 NaRpoTm1 and NaRpoTm2

To determine the subcellular localization of the *NaRpoTm1*- and *NaRpoTm2*-encoded polypeptides, putative N-terminal transit peptides were fused to the GFP coding region and the constructs were used for transformation of *Arabidopsis* protoplasts. The transformed protoplasts were examined by confocal laser scanning microscopy.

Assuming that translation starts from the first encoded methionine, the following constructs were generated: *Na-RpoTm1*$_{met}$*-GFP* and *Na-RpoTm2*$_{met}$*-GFP* with the first encoded methionine cloned

immediately downstream of the 35S promoter for forced translation, *Na-RpoTm1$_{utr}$-GFP* and *Na-RpoTm2$_{utr}$-GFP* containing the whole 5'untranslated region, and *Na-RpoTm1$_{mut}$-GFP* and *Na-RpoTm2$_{mut}$-GFP*, in which the methionine had been mutated to isoleucine (see Fig. 14 and Fig. 15).

Both Na-RpoTm1-GFP and Na-RpoTm2-GFP constructs exhibited the same characteristic subcellular localization (Fig. 34): In the case of Na-RpoTm1$_{met}$-GFP (Fig. 34D) and Na-RpoTm2$_{met}$-GFP (Fig. 34G), with forced translation from the first encoded methionine, GFP fluorescence was observed only in mitochondria. The constructs containing the full-length of the 5' untranslated leader sequence, Na-RpoTm1$_{utr}$-GFP (Fig. 34E) and Na-RpoTm2$_{utr}$-GFP (Fig. 34H) showed also exclusive mitochondrial targeting. When the mutated full-length transit peptides Na-RpoTm1$_{mut}$ (Fig. 34F) and Na-RpoTm2$_{mut}$ (Fig. 34I) (methionine substituted by isoleucine) were used, GFP fluorescence was detectable neither in mitochondria, nor in chloroplasts. It was concluded that the AUG at position +177 (NaRpoTm1) and +253 (NaRpoTm2), respectively, is the only available *RpoT* start codon, from which translation of a polypeptide with mitochondrial targeting properties is initiated.

3.4.2.3.2 NaRpoTp

In the full length cDNA of *NaRpoTp* no in-frame AUG start codon could be detected in the 5' coding region. Instead, a CUG codon was found at position +148, from which translation could be initiated. Initiation of translation from non-AUG codons is not common in plants and has been described only for a few genes (Gordon *et al.*, 1992). Interestingly, one of the *RpoT* genes in tobacco, coding for a plastid-localized RNA polymerase, is also initiated from a CUG codon (Hedtke *et al.*, 2002). To investigate whether the *NaRpoTp* gene could be initiated from the CUG (leucine) at postion +148, the following three Na-RpoTp-GFP constructs were generated: Na-RpoTp$_{leu}$-GFP, with the first encoded leucine (+148) cloned immediately downstream of the 35S promoter for forced translation; Na-RpoTp$_{utr}$-GFP containing the whole 5'untranslated region of 236 nt; and Na-RpoTp$_{mut}$-GFP, where the encoded leucine had been substituted by threonine (see Fig. 16). The Na-RpoTp$_{leu}$-GFP construct gave rise to bright-green GFP fluorescence in chloroplasts (Fig. 34J, green channel) which overlapped with the red chlorophyll autofluorescence, clearly confirming co-localization of red and green fluorescence in chloroplasts (Fig. 34J, merged image). The localization assay of the Na-RpoTp$_{utr}$-GFP construct also resulted in clear co-localization of the red chlorophyll autofluorescence and the green GFP fluorescence in the chloroplasts (Fig. 34K), whereas expression of *Na-RpoTp$_{mut}$-GFP* (Fig. 34L) completely abolished

import of the GFP to the chloroplasts. These data provide convincing evidence that translation of *NaRpoTp* is solely initiated from the CUG codon at position +148.

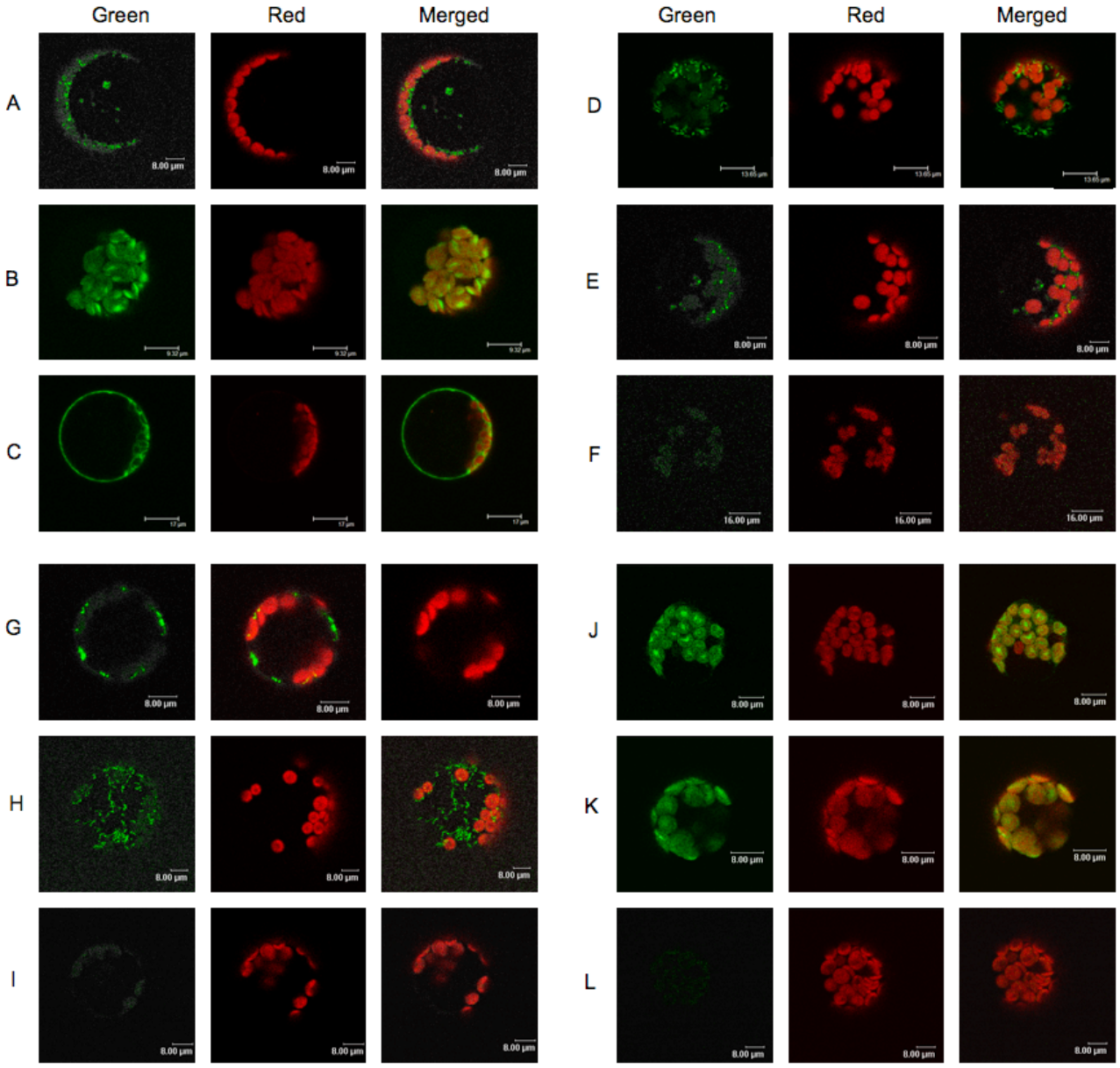


Fig. 34 Subcellular localization of the *NaRpoT* gene products observed by confocal laser scanning microscopy of transformed *Arabidopsis* protoplasts. The images depict fluorescence patterns (green, red and merged channels) of control constructs targeting GFP to mitochondria (A), plastids (B), vector control containing no transit peptide (C), Na-RpoTm1met-GFP (D), Na-RpoTm1utr-GFP (E), Na-RpoTm1mut-GFP (F), Na-RpoTm2met-GFP (G), Na-RpoTm2utr-GFP (H), Na-RpoTm2mut-GFP (I), Na-RpoTpleu-GFP (J), Na-RpoTputr-GFP (K) and Na-RpoTpmut-GFP (L). Scale bars show the size.

3.5 Phylogenetic analysis

When the amino acid sequences of the RpoT polymerases from *Selaginella moellendorffii* and *Nuphar advena* were aligned with the sequences of the three RpoT polymerases from *Arabidopsis* and the three enzymes from *Physcomitrella*, high similarity was observed (Fig. 27). According to their coding sequences and remarkable conservation of intron numbers and positions (black triangles in Fig. 27), the *SmRpoT* as well as the *NaRpoT* genes most likely result from the same common ancestor giving rise to both the moss and angiosperm *RpoTs*.

Based on a multiple alignment of 41 RpoT sequences (see "Material and Methods"), phylogenetic trees were reconstructed by using programs Tree-Puzzle and Mr. Bayes 3.1, respectively, to elucidate molecular phylogeny of the RpoT polymerases and to determine the evolutionary position of the polymerases identified and described in the present study. Both programs (Tree-Puzzle and Mr. Bayes 3.1) resulted in essentially the same topology, with the trees shown in Fig. 35 and Fig. 36, respectively. In both consensus trees, angiosperm RpoT polymerases cluster into two clearly discernible groups, which might be sub-divided into sub-groups: one consisting of plastid-localized polymerases, and the other of mitochondrial-localized and dual-targeted polypeptides. The RpoT polymerases from the moss *Physcomitrella* as well as the single enzyme from the spikemoss *Selaginella* form sister groups to the angiosperm polymerases. Interestingly, the position of the *Selaginella* RpoT is in conflict with the phylogenetic position of the species: *Selaginella* is, compared to *Physcomitrella,* the later branching species (compare to Fig. 7 in the "**Introduction**" part). Thus the spikemoss polymerase seems to be closer related to the plant ancestor RpoT sequence than those of the moss (for details, see "**Discussion**"). Whereas SmRpoT, like the moss polymerases, does not cluster with the branches of well separated plastid and mitochondrial (and dual targeted) polymerases, building a sister group to all angiosperm RpoTs, the RpoT polymerases from the basal eudicot *Nuphar advena* cluster with the branches of plastid and mitochondrial/dual targeted sequences: NaRpoTm1 and NaRpoTm2 within the mitochondrial, and NaRpoTp within the plastid branch. This is also an indication of the appearance of a NEP activity (NaRpoTp) among the basal eudicots.

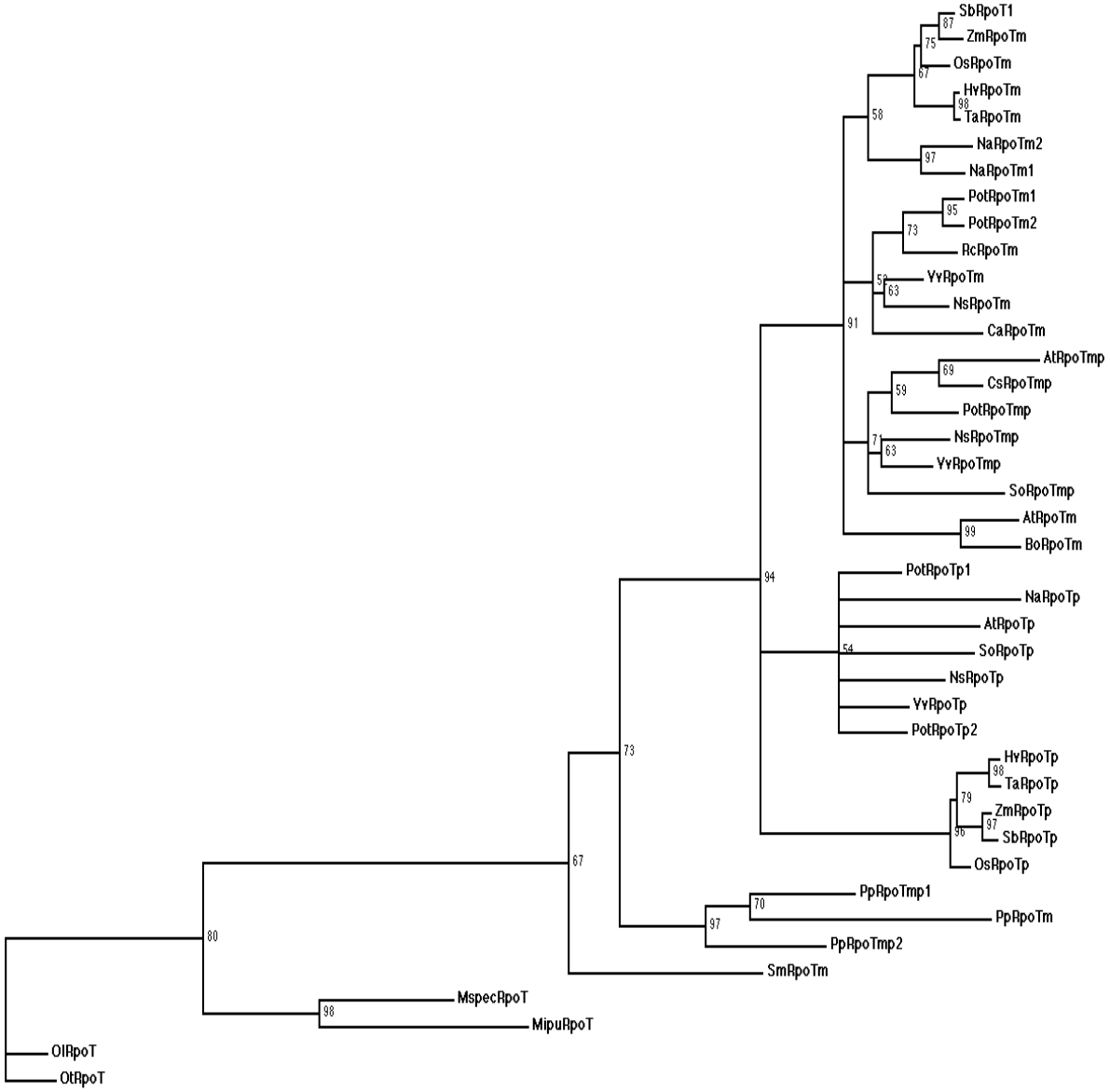


Fig. 35 Phylogenetic analysis of RpoT sequences. Tree of plant RpoT protein sequences was constructed by using program Tree-Puzzle, based on an alignment of conserved blocks (see 'Materials and Methods' part.). Sequences and their accession numbers (m designates mitochondrial enzyme, p plastid enzyme, and mp enzyme with dual targeting): AtRpoTm—*Arabidopsis thaliana* (P92969); AtRpoTp—*A. thaliana* (O24600); AtRpoTmp—*A. thaliana* (CAC17120); BoRpoTm—*Brassica oleracea var. alboglabra* (XP_002308414.1), CaRpoTm—*Chenopodium album* (CAA69305); CsRpoTmp—*Cleome spinosa* (DQ415921); HvRpoTm—*Hordeum vulgare* (AJ586899); HvRpoTp—*H. vulgare* (AJ507396); MipuRpoT—*Micromonas pusilla* (EEH56417.1); MspecRpoT—*Micromonas sp. RCC299* (XP_002503703.1); NaRpoTm1—*Nuphar advena* (FN811768), NaRpoTm2—*N. advena* (FN820498), NaRpoTp—*N. advena* (FN811769); NsRpoTm—*Nicotiana sylvestris* (AJ416568); NsRpoTp—*N. sylvestris* (AJ302020); NsRpoTmp—*N. sylvestris* (AJ302019); OsRpoTm—*Oryza sativa* (AB096014); OsRpoTp—*O.sativa* (AB096015); OtRpoT—*Ostreococcus tauri* (CAL55557.1); PotRpoTm1, PotRpoTm2, PotRpoTmp, PotRpoTp1, PotRpoTp2—*Populus trichocarpa*, sequences derived from five RpoT polymerase genes identified in the genome of poplar (LG_X:13144158-13137666, scaffold_165:313898-322951, LG_XVIII:10512245-10519982, see http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html; U. Richter, unpublished data); PpRpoTmp1—*Physcomitrella patens* (CAC95163); PpRpoTmp2—*P. patens* (CAC95164); PpRpoTm—*P. patens* (see Legend to Figure 5); SbRpoTm—*Sorghum bicolor* (XM_002460990), SbRpoTp—*S. bicolor* (XM_002437329); SmRpoT— *Selaginella moellendorfii* (CAP70041); SoRpoTp—*Spinacia oleracea* (Y18853); SoRpoTmp—*S. oleracea* (Y18852); TaRpoTm—*Triticum aestivum* (AAF32492); TaRpoTp—*T. aestivum* (AAB01085); VvRpoTm—*Vitis vinifera* (AM483136), VvRpoTp—*V. vinifera* (AM453066), VvRpoTmp—*V. vinifera* (AM488491); and ZmRpoTm—*Zea mays* (AAD22977); ZmRpoTp—*Z. mays*

(AAD22976). The accession numbers of OIRpoT from *Ostreococcus lucimarinus* and RcRpoTm from *Ricinus communis*, do not exist in the database. These two sequences were extracted from the corresponding genome raw data together with the EST data, and then analyzed.

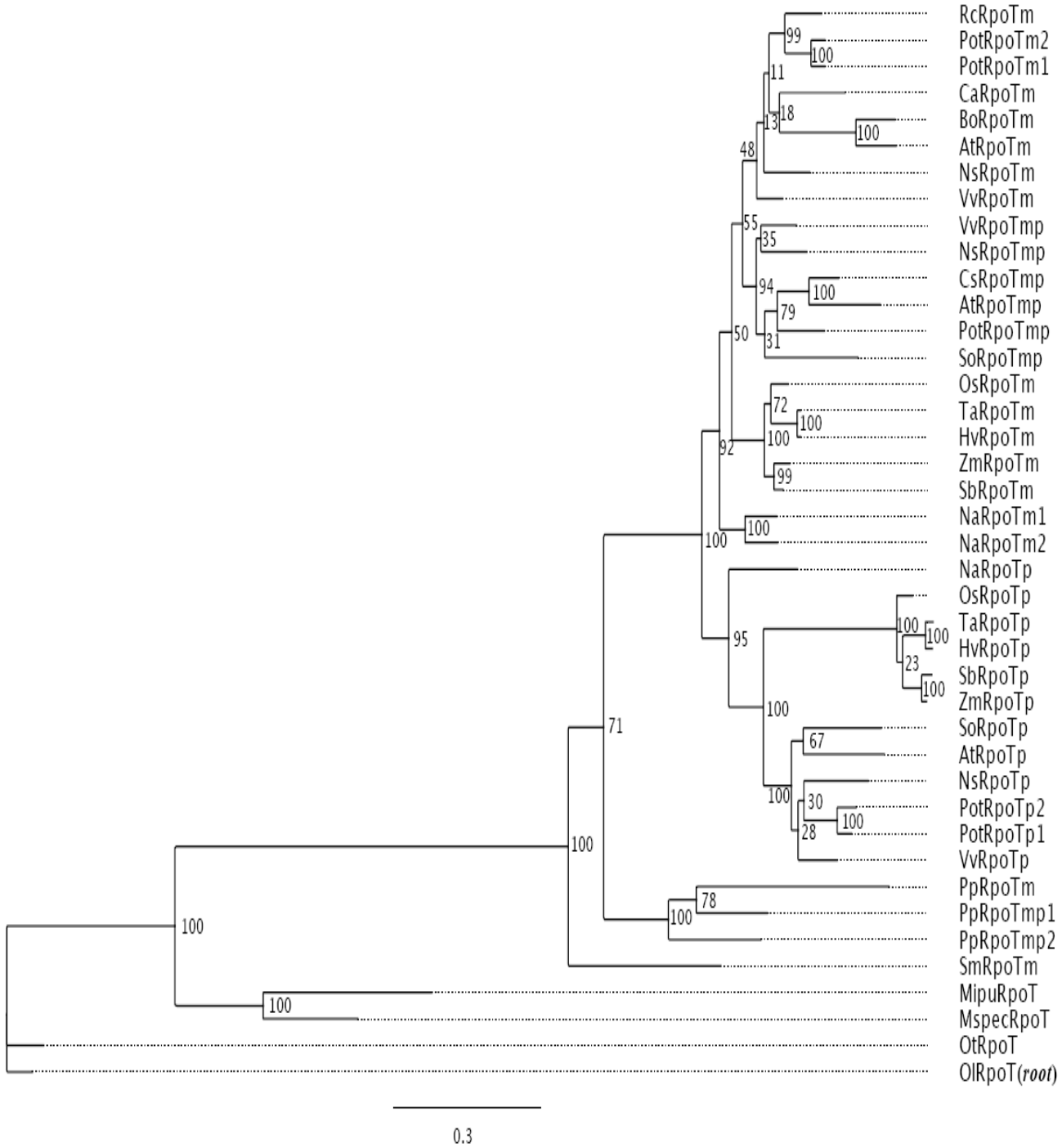


Fig. 36. Phylogenetic analysis of RpoT sequences. ML (Bayesian) tree of plant RpoT protein sequences was reconstructed based on an alignment of conserved blocks (see 'Materials and Methods' part.). For sequences and their accession numbers (m designates mitochondrial enzyme, p plastid enzyme, and mp enzyme with dual targeting), please refer to the legend to Fig. 35.

4 Discussion

4.1 General properties of *RpoT* genes from *Selaginella* and *Nuphar*

Genes encoding phage-type mitochondrial and plastid RNA polymerases have been identified from numerous angiosperm species (Weihe, 2004). In contrast, regarding the RpoT polymerases of lower plants, the knowledge is so far only limited to the moss *Physcomitrella patens* (Richter *et al.*, 2002; Kabeya *et al.*, 2002). Here we show that the spikemoss *Selaginella moellendorffii* and the water lily *Nuphar advena* (a basal eudicot), encode one and three RpoT polymerases, respectively.

We were able to clone the *RpoT* genes from both *S. moellendorffii* and *N. advena,* and characterized the properties of the encoded proteins. The SmRpoT (1,002 amino acids) and NaRpoT proteins (996, 990, and 985 amino acids, respectively) exhibit the characteristic domains that are highly conserved between all RpoT polymerases, including the residues shown to be essential and located within the catalytic pocket of the polymerase (D537, K631, Y639, G640, D812, numbers as given for T7 RNA polymerase). Moreover, the gene structure of both the single *Selaginella RpoT* and the three *Nuphar RpoTs* is highly similar: all 18 intron positions are conserved with those of all angiosperm *RpoT* genes characterized thus far, indicating that this feature had been retained since at least 140-150 million years ago (late Jurassic-early Cretaceous), when the monocots and eudicots diverged (Chaw *et al.*, 2004), supporting the hypothesis that a common ancestral gene gave rise to all land plant *RpoT* genes.

4.2 Initiation of translation of *NaRpoTp* occurs at a non-canonical CUG codon

Phylogenetic analysis (Hess & Borner, 1999) indicated that tobacco *RpoTp* encodes a plastid phage-type RNA polymerase. However, the first ATG triplet of the *RpoTp* reading frame did neither in predictions nor in experiments display plastid targeting properties. Closer inspection of 5' sequences revealed a potential non-AUG initiation codon (CUG). Experiments with GFP fusion constructs revealed that the tobacco *RpoTp* is initiated at a non-canonical CUG codon (Hedtke *et al.*, 2002). Thus, *Nicotiana RpoTp* is one of the rare non-viral plant genes initiating translation exclusively at a non-AUG codon, which resides within an optimal sequence context (Joshi *et al.*, 1997). Although the percentage of non-AUG translation starts in plants is low, the *RpoTp* example underlines the importance of a careful assessment of sequence data, especially if N-terminal targeting sequences are to be predicted. We could show that a non-canonical initiation of translation also occurs in the *RpoTp* gene of *Nuphar*. In the 5' part of the *N. advena* cDNA, no canonical start codon was identified, with the first ATG triplet occurring only at position 412. However, a potential non-AUG initiation codon (CUG) was revealed at position 148. Translation from this codon would

yield a sequence of clearly plastid targeting properties, as predicted by two prediction algorithms (TargetP and Predotar). Three different GFP fusions were designed to test the translation initiation capacity of this CUG codon. The results proved a plastid import of the derived amino-terminus (Fig. 34J), as well as an efficient translation initiation at the CUG within the context of the full-length 5'-UTR (Fig. 34K) that could be abolished by modifying the codon to CAC (Fig. 34L). Thus, the *N. advena RpoTp* belongs to the rare non-viral plant genes (Riechmann *et al.*, 1999) that initiate translation exclusively at a non-AUG codon. This is the second case of non-AUG translation initiation among *RpoT* genes specifying plastid-localized RNA polymerases following the tobacco *RpoTp* gene which also starts from a CUG codon (Hedtke *et al.*, 2002).It is tempting to speculate that, this sequence feature appeared early in evolution, within basal angiosperms, and has been conserved in some, but not all higher plant *RpoTp* genes.

4.3 Subcellular localization of the RpoTs

4.3.1 *Selaginella moellendorffii*

The N-terminus of SmRpoT shows properties of a mitochondrial transit peptide and *in silico* analysis using two prediction algorithms (TargetP and Predotar) indeed showed mitochondrial targeting with significant scores. Previous studies showed different targeting properties of the *Physcomitrella* RpoT transit peptides if assayed with or without the flanking 5'-UTR because of the usage of different start codons (Kabeya & Sato, 2005). Therefore, we included in our study constructs containing the full 5' UTR context of the translation initiation sites. By applying translational fusions of the putative SmRpoT transit peptide with GFP to the protoplasts from *Arabidopsis* and *Selaginella,* respectively, we demonstrated that this transit peptide confered exclusively mitochondrial import, regardless of whether translation was forces from the first encoded methionine or from constructs containing the whole 5' UTR. Subcellular localization was analyzed by transient expression in both *Selaginella* and *Arabidopsis* protoplasts, since we could not exclude that the organelles' import apparatus of *Arabidopsis* might not be compatible with the transitpeptides of the spikemoss construct. Our work is the first demonstration of usage of protoplasts from *Selaginella* for transient gene expression studies. The protoplasts from *S. moellendorfii* are smaller in size, more difficult in handling and observation, and prone of early death, compared with those from *Arabidopsis.* The protoplasts contain fewer chloroplasts with a lower chlorophyll autofluorescence, possibly caused by lower pigment content in the organelles, which resulted in lower signals (right panel in Fig. 33). Nevertheless, we managed to observe GFP expression in the protoplasts. Mitochondrial import of SmRpoT-GFP was demonstrated for both construct with forced translation from the first encoded methionine and construct containing the full-length 5'-UTR of the *SmRpoT* gene. Moreover, replacement of the potential start codon by an

isoleucin codon fully abolished transport into organelles. Essentially, the same results were obtained when the transient expression of the constructs was followed in *Arabidopsis* protoplasts. This indicates that the import machinery of *Arabidopsis* (and probably other angiosperms which are often used for transient expression studies, like tobacco) is compatible with transit peptides from lower plants, in this case, a lycophyte. The import studies clearly showed that *SmRpoT* encodes the mitochondrial RNA polymerase of *S. moellendorfii*. No indication for any dual targeting, i.e. additional import into chloroplasts (as ocurring in the case of RpoTmp polymerases of eudicots and in *Physcomitrella*), was found.

4.3.2 *Nuphar advena*

To apply a homologous system for transient expression studies of GFP fusion constructs, we tried to isolated protoplasts from *N. advena* (results not shown). The same protocol for isolation of protoplasts from *Selaginella moellendorffii* (see 2.20.2.) was applied to *Nuphar advena*. Under the optical microscope, most protoplasts were observed alive, and the yield of protoplast was around 2 x 10^5 per gram of leaves, on a similar level as for *Arabidopsis* and *Selaginella*. Although the cells were more fragile than the protoplasts from *Arabidopsis*, in general, protoplast isolation from *Nuphar advena* was successful. As well, the same PEG transformation procedure (see 2.20.3.) was applied to transform *Nuphar advena* protoplasts. Unfortunately, after transformation most of the cells died, with only about 10% of cells still alive (compared to more than 30% for *Arabidopsis* or *Selaginella* protoplasts). There might be two possible reasons: first, the protoplasts of *Nuphar advena* are so sensitive that they could not survive long enough till completion of PEG transformation; second, trace PEG might be toxic to the protoplasts from *Nuphar advena*, as PEG can destroy cell membrane. As a consequence, targeting studies of Nuphar GFP-RpoT fusion constructs were performed only with *Arabidopsis* protoplasts (see Fig. 34 in Section 3.4.2.3).

In the protoplast system of *Arabidopsis*, the N-termini of two of the three RpoTs of *N. advena* (NaRpoTm1 and NaRpoTm2) showed properties of mitochondrial transit peptides. Using translational fusions of the putative NaRpoT transit peptides with GFP, we demonstrated that these transit peptides confer exclusively mitochondrial import, and mitochondrial import of NaRpoTm1- and NaRpoTm2-GFP was also maintained when the fusion constructs contained the full-length 5'-UTRs of the genes (Fig. 34). Thus, we conclude that *Nuphar advena* encodes two phage-type mitochondrial RNA polymerases. As discussed above, translation initiation from the third RpoT from *N. advena*, NaRpoTp, is initiated from a CUG codon, and the translational GFP fusion with the transit peptide initiated from this codon clearly showed plastid targeting properties (see Section 3.4.2.3.2.). Like in the case of the *Selaginella* RpoT, none of the *N. advena* RpoTs posses dual

targeting properties. The subcellualr localization of the three *N. advena* RpoT polymerases was confirmed by their positions in phylogenetic trees reconstructed from 41 RpoT sequences.

4.4 The lycophyte *Selaginella moellendorffii* contains a single *RpoT* gene encoding a mitochondrial RNA polymerase

The spikemoss *Selaginella* belongs to the lycophytes. That there is only one *RpoT* gene identified from *Selaginella moellendorfii* in this study is a surprising observation because the nuclear genome of *Selaginella* does not appear to encode any other *RpoT*s, based on the DNA filter hybridization shown in Fig. 21C, using low stringency conditions with a probe from the highly conserved 3' region. Screening of the database of the *Selaginella* nuclear genome project (http://www.phytozome.net/selaginella.php) using the BlastX algorithm and different query sequences from the most conserved parts of plant *RpoT* genes did not yield any other hits either. Using this method, *RpoT* sequences from distantly related species (such as fungi) or from species with an extreme codon use (for example, *Chlamydomonas*) can be identified. Thus, the apparent presence of only one *RpoT* gene in *Selaginella* is in striking contrast, not only to the small *RpoT* gene families found in angiosperms, but also to the situation in the only deep-branching land plant examined thus far, the moss *P. patens*, which contains three *RpoT* genes (Richter *et al.*, 2002), (http://genome.jgi-psf.org/phypa1_1/phypa1_1.home.html). Thus, lycophytes seem not to contain a plastid-localized phage-type RNA polymerase, suggesting that NEP, i.e. nuclear encoded plastid RNA polymerases, occur at a later stage of the green lineage of evolution.

4.5 *Nuphar advena* encodes three phage-type RNA polymerases

In the basal angiosperm *N. advena* three *RpoT* genes were identified. Identification was based on screening of a BAC library which revealed three different *RpoT* genes. The library used for screening had a redundancy of approximately 7, i.e. contained seven genomic equivalents. The twenty-four positive BAC clones were identified using an *RpoT* cDNA fragment from *Selaginella* as probe, under non-stringent conditions. DNA fingerprinting of the BACs suggested that they represented three similar, yet individual genes. Taking into account the redundancy of the library, one would expect approximately the number of positive clones as we have identified. Would there be a fourth gene, positive clones with a comparatively higher number (up to 28) should have been detected. To check the gene copy number, we performed Southern blot analyses using a mixed probe containing conserved 3' part cDNA fragments from all three *RpoT* genes to account for a situation when a hypothetical fourth gene would be more similar to one of the three genes. Southern hybridization with *N. advena* DNA is a challenge because of the enormous genome size of *Nuphar*. Nevertheless, we identified in the blots bands of predicted sizes corresponding to the three

identified genes. In some digests, very faint additional bands were observes (see Fig. 32). However, we suggest that these bands rather originate from incomplete digestion or cross-hybridization to similar, non-*RpoT* sequences than from a fourth *RpoT* gene. In the latter case, we would expect the bands to be more prominent and to occur in all digests. In summary, our experiments indicated that the *RpoT* gene family of *N. advena* consists of three genes encoding two mitochondrial and one plastid RNA polymerase.

4.6 Molecular phylogeny of the *RpoT* gene family

In the phylogenetic trees of plants (see Fig. 37), lycophytes form a sister group to the euphyllophytes, which comprise seed plants and monilophytes. *Physcomitrella*, at a deeper branch, is likely in a sister position to hornworths, lycophytes, and all other tracheophytes. The nuclear genome content of the lycophyte *Selaginella moellendorffii* has the smallest genome size reported thus far for a land plant species, about two-thirds that of *Arabidopsis thaliana*. The whole genome sequencing of *Selaginella moellendorffii* was recently completed by the Department of Energy Joint Genome Institute. As a clade intermediate between nonvascular plants and all other vascular plants, *Selaginella moellendorffii* might serve as a bridge for evolutionary comparisons.

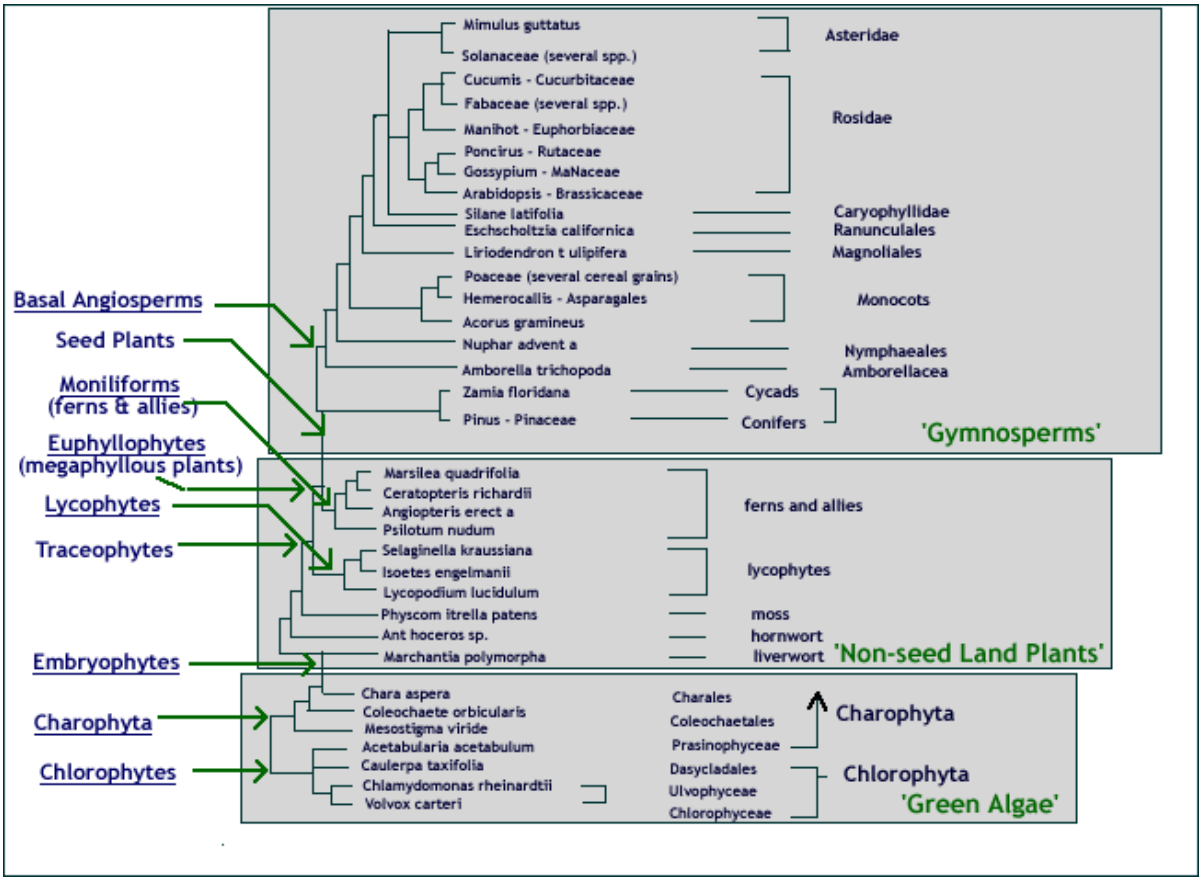


Fig. 37 A comprehensive phylogenetic tree of plants (from http://www.greenbac.org/ tree.html).

The existence of three *RpoT* genes in *Physcomitrella* and only one in *Selaginella* suggests that the spikemoss has lost all gene copies except one or that the last common ancestor of bryophytes and lycopods originally contained only one gene in contrast to *Physcomitrella*. The phylogenetic tree of RpoT (see Fig. 35 and Fig. 36 in **Results** part) shows that duplications in *Physcomitrella* and in the angiosperms are most likely different phylogenetic events. Thus, duplications in angiosperms occurred after splitting of the lycophytes because *Selaginella* encodes only one *RpoT* gene. The fact that the single *RpoT* gene in *Selaginella* encodes a mitochondrial RNA polymerase supports the idea that the ancestral *RpoT* gene of all land plants originally encoded a mitochondrial-localized polymerase. Our data demonstrate that land plants do not necessarily contain more than one phage-type RNA polymerase and may exist without a NEP activity.

Although there are three phage-type RNA polymerases in *Physcomitrella* (one mitochondrial and two dual-targeted) and eudicots (one mitochondrial, one plastid and one dual-targeted), the localization of the three *Nuphar* RpoT polymerases shows a new pattern. The three *N. advena RpoT* genes are suggested to result from gene duplication(s) from the original gene encoding a mitochondrial RNAP, based on the following observations: 1) The coding regions of the three *RpoT* genes are highly similar with regard to the gene sequence (shown in Tab. 14); 2) All 18 introns are located at identical positions among these three genes (shown in Fig. 26); 3) The deduced amino acid sequences of the three *RpoT* genes are highly similar (shown in Fig. 27). Both mono- and eudicotyledonous plants possess a solely plastid localized phage-type RNA polymerase (RpoTp) together with a purely mitochondrial localized RpoT enzyme (RpoTm) and, in the case of eudicots, a third phage-type polymerase with dual localization in both organelles is found. Our data suggested that all RpoTp proteins are orthologous, or in other words, decent from a common duplication event that took place in a common ancestor of all flowering plants. Thus far it is unknown whether ferns or gymnosperms contain nuclear genes encoding plastid localized phage-type RNA polymerases as well. Since the duplication event giving rise to the second NEP activity in eudicots is clearly more recent, identification of a purely plastid localized phage-type RNAP in the basal eudicot *Nuphar advena*, orthologous to all other RpoTp enzymes of flowering plants, suggests that the acquisition of a nuclear encoded transcriptional activity for plastids, not present in lycopods, took place after the split of lycopods from all other tracheophytes, with or before the rise of flowering plants. Moreover, the lack of a dual-targeted RpoTmp both in *Nuphar* and in monocots suggests that the RpoTmp enzyme detected in eudicots is an 'invention' due to an *RpoTm* gene duplication that might have occurred only after the separation of monocots and eudicots. It is therefore that the putative plastid targeting sequences as present in two of the three *Physcomitrella* RpoT proteins are clearly species- or lineage-specific convergent inventions. It is interesting that multiple mitochondrial RNA

polymerases as found in *Physcomitrella* and eudicots were indentified in *Nuphar* as well. The fixation of duplicated *RpoT* genes led to convergent multiplicity of mitochondrial RNA polymerases in *Nuphar*, *Physcomitrella* and eudicots, not found in any other eukaryotic lineage. Further investigations have to show if there exits a division of labor between the two mitochondrial RNA polymerases in *Nuphar* similar to that found in *Arabidopsis* (Kuhn *et al.*, 2009) and how NEP and PEP share activities during the transcription in *Nuphar* chloroplasts.

The evolutionary trends of components of plant organelle transcription observed and described thus far are based on too few model plants to derive a complete scenario of molecular phylogeny. We have to investigate a greater diversity of plants along the phylogenetic tree to understand the evolution of the transcriptional machineries in plants and, within them, the functional importance of phage-type RNA polymerases.

4.7 *RpoT* genes in other deep branching plant species and algae

Duplications of *RpoT* genes have taken place several times during the process of plant evolution. Since *Chlamydomonas* possess only one *RpoT* gene, gene duplication may have occurred not before the evolution of plants with specialized tissues. Therefore, besides *Selaginella moellendorffii* and *Nuphar advena*, we tried to isolate *RpoT* genes from more deep branching plant species and algae (results not shown). Thus, during the initial phase of this project, identification of *RpoT* genes from *Coleochaete orbicularis* (green algae), *Ginkgo biloba*, *Adiantum capillus-veneris, Ambrella* and *Marchantia polymorpha* was attempted by amplification of *RpoT* sequences using degenerated primers derived from angiosperm species. A multitude of species was used, because we anticipated that the needed quality of the DNA preparations and/or successful amplification will be not reached with all species. Due to various reasons, isolation of *RpoT* genes from these species failed or was not completed, with some results discussed in detail below.

4.7.1 *Coleochaete orbicularis*

A BAC filter of *Coleochaete orbicularis* (Genome size: 94Mb) was obtained from the Green BAC Project, with an average insert size of 160 kb, covering 31 genome equivalents. Using the 1.5 kb cDNA probe from *Selaginella moellendorffii*, 8 positive (although weak-hybridizing) clones were identified. None of the identified clones was related to RpoT sequences, as determined by sequencing of identified subclones. We could only speculate that misidentification of the clones might result from a highly deviating sequence of the *RpoT* gene in this alga, or by bad quality of the filters.

4.7.2 *Marchantia polymorpha*

Southern hybridization of a *Marchantia* BAC filter obtained from the Green BAC Project didn't reveal any clear positive clones due to unknown reasons. As both *Marchantia polymorpha* and *Selaginella moellendorffi* belong to non-seed land plants, we decided to concentrate on *Selaginella moellendorffii*, which was successful.

4.7.3 *Ginkgo biloba*

Ginkgo is a living fossil, dating back 270 million years. *Ginkgo biloba* as a unique species with no close living relatives would thus be a very interesting species for the study of *RpoT* genes. Our strategy to identify *RpoT* sequences using PCR with degenerated primers derived from angiosperm sequences failed. As in the above cases, we suggested that the potential *RpoT* gene(s) in *Ginkgo* considerably deviate from those in the higher plant sequences (which were the basis of primer design).

4.7.4 *Adiantum capillus-veneris*

Adiantum capillus-veneris is a species of fern in the genus *Adiantum* with a subcosmopolitan worldwide distribution. The analysis of *RpoT* gene from *Adiantum capillus-veneris* followed the strategy applied to *Selaginella moellendorffii* and *Nuphar advena,* and is still in progress. The available results suggest that there is only one *RpoT* gene in *Adiantum*. 2.5 kb of the cDNA sequence of the *Adiantum RpoT* gene (most part of the CDS) were sequenced (see Appendix 6.2.). The 5'- and 3'- are not completed yet, and therefore no information on subcellular localization is available.

4.8 Concluding marks

Initial attempts to amplify *RpoT* sequences from deep branching plants and algae widely failed. For PCR amplification by using both genomic DNA and cDNA from around 30 plant species, different thermo-stable DNA polymerases as well as amplification protocols were tried to overcome the difficulties resulting obviously from deviating *RpoT* target sequences. In none of any DNA/cDNA-primer combinations could *RpoT*-specific fragments be amplified from most of the species which we examined. As this failure in amplification occurred with different species and primer pairs, we concluded that the *RpoT* target sequences in most of the lower plants and algae deviate from the higher plant sequences (which were the basis of primer design), which made it impossible to obtain *RpoT* sequences with this strategy.

Due to the limited time frame, only two species (*S. moellendorffi* and *N. advena)* that according to their positions in phylogenetic trees promised to give the most conclusive data about *RpoT* gene evolution were selected for in-depth investigation in this study. We prepared genomic DNA and cDNA from these two plants and were able to amplify partial "tag" *RpoT* sequences, benefiting from the availability of partial *RpoT* raw sequence data in the NCBI database in the case of *Selaginella*. Primers were designed for PCR amplification, RACE and sequencing. The resulting sequences were aligned, followed by reconstruction of phylogenetic trees. Southern hybridizations were performed to check the *RpoT* gene number in *S. moellendorffi* and *N. advena*. The coding regions for the putative N-terminal transit peptides of the gene products were fused to a reporter gene (GFP), and the constructs were used for transfection of protoplasts to determine the subcellular localization of the RpoT polymerases by confocal laser scanning microscope.

Finally, we were able to show that the spikemoss *Selaginella moellendorfii* and the waterlily *Nuphar advena*, encode one (targeting to mitochondria) and three RpoT polymerases (two targeting to mitochondria and the third one targeting to chloroplast), respectively. Based on the data outlined above, it is evident that duplications of genes for phage-type RNA polymerases have taken place several times during plant evolution.

We tried to find out when during plant evolution for the first time a plastid phage-type polymerase occurred, i.e. in which group of plants the two gene duplication events have taken place that gave rise to the plastid RNA polymerase observed in monocots and eudicots and to the dually targeted RNA polymerase observed only in eudicots. This information is not only important for the understanding of the evolutionary pathway leading to the integration of plastids into the eukaryotic cell, but for the correlation of the occurrence of the investigated genes with a certain complexity of the plant (single cellular vs. increasing complexity of tissues and organs) which may also contribute to the understanding of the functional roles these polymerases play within the plant cell: why do plastids need phage-type polymerases (NEP) in addition to the bacterial-type polymerases (PEP). Our results indicate that a plastid phage-type is already present in the basal angiosperm *N. advena*, and probably in other basal angiosperms.

Since *Chlamydomonas* has only one *RpoT* gene, gene duplication may have occurred not before the evolution of plants with differentiated tissues. *Chara* and related algae could be interesting subjects in this context as they are multicellular and develop different types of tissues. Furthermore, they are among those algae that most likely served as ancestors of all land plants. Further interesting taxa in the context of the evolution of RpoT polymerases should be selected from major branches representing monophyletic clades of land plants, i.e. representatives of the *Psilotales, Cycadopsida,*

Ginkgo and *Coniferopsida*. Data from these plants are expected to provide further insight into the frequency of occurrence of independent gene duplications among land plants and answer the question whether the first gene duplication leading to a plastid RpoT polymerase in monocots and eudicots happened before the evolution of angiosperms. Several more angiosperms could be selected to clarify when the first (if not already identified within the taxa mentioned above) and the second gene duplication (until now only known from the eudicots) occurred. Focus should be put on such angiosperms as *Ambrella*, that are proposed to have appeared early in angiosperm evolution, and on monocots other than the cereals, since thus far among monocots *RpoT* genes have been characterized only in cereals. Sequencing nuclear *RpoT* genes from more species including algae will broaden our knowledge about *RpoT* evolution and answer the question when and how often during the evolution of 'green eukaryotes' was the gene for the mitochondrial RNAP duplicated, permitting evolution of a gene for a plastid enzyme.

5 References

Adams KL, Cronn R, Percifield R, Wendel JF. **2003.** Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences of the United States of America* **100:** 4649-4654.

Akashi K, Grandjean O, Small I. **1998.** Potential dual targeting of an Arabidopsis archaebacterial-like histidyl-tRNA synthetase to mitochondria and chloroplasts. *Febs Letters* **431:** 39-44.

Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, de Pouplana LR, Martinez-Castilla L, Yanofsky MF. **2000.** An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 5328-5333.

Ammiraju JS, Luo M, Goicoechea JL, Wang W, Kudrna D, Mueller C, Talag J, Kim H, Sisneros NB, Blackmon B, Fang E, Tomkins JB, Brar D, MacKill D, McCouch S, Kurata N, Lambert G, Galbraith DW, Arumuganathan K, Rao K, Walling JG, Gill N, Yu Y, SanMiguel P, Soderlund C, Jackson S, Wing RA. **2006.** The Oryza bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus Oryza. *Genome Res.* **16:** 140-147.

Baba K, Schmidt J, Espinosa-Ruiz A, Villarejo A, Shiina T, Gardestrom P, Sane AP, Bhalerao RP. **2004.** Organellar gene transcription and early seedling development are affected in the rpoT;2 mutant of Arabidopsis. *Plant Journal* **38:** 38-48.

Banks JA. **2009.** Selaginella and 400 Million Years of Separation. *Annual Review of Plant Biology* **60:** 223-238.

Binder S, Brennicke A. **2003.** Gene expression in plant mitochondria: transcriptional and post-transcriptional control. *Philosophical Transactions of the Royal Society B-Biological Sciences* **358:** 181-188.

Binder S, Hatzack F, Brennicke A. **1995.** A Novel Pea Mitochondrial In-Vitro Transcription System Recognizes Homologous and Heterologous Messenger-Rna and Transfer-Rna Promoters. *Journal of Biological Chemistry* **270:** 22182-22189.

Bohne A-V, Weihe A, Borner T. **2009.** Transfer RNAs inhibit Arabidopsis phage-type RNA polymerases. *Endocytobiosis Cell Research.* **19:** 63-69.

Cahoon AB, Harris FM, Stern DB. **2004.** Analysis of developing maize plastids reveals two mRNA stability classes correlating with RNA polymerase type. *Embo Reports* **5:** 801-806.

Caoile AGFS, Stern DB. **1997.** A conserved core element is functionally important for maize mitochondrial promoter activity in vitro. *Nucleic Acids Research* **25:** 4055-4060.

Cermakian N, Ikeda TM, Cedergren R, Gray MW. **1996.** Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucleic Acids Research* **24:** 648-654.

Cermakian N, Ikeda TM, Miramontes P, Lang BF, Gray MW, Cedergren R. **1997.** On the evolution of the single-subunit RNA polymerases. *Journal of Molecular Evolution* **45:** 671-681.

Chang CC, Sheen J, Bligny M, Niwa Y, Lerbs-Mache S, Stern DB. **1999.** Functional analysis of two maize cDNAs encoding T7-like RNA polymerases. *Plant Cell* **11:** 911-926.

Chaw SM, Chang CC, Chen HL, Li WH. **2004.** Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J.Mol.Evol.* **58:** 424-441.

Cheetham GMT, Jeruzalmi D, Steitz TA. **1999.** Structural basis for initiation of transcription from an RNA polymerase-promoter complex (vol 399, pg 80, 1999). *Nature* **400:** 89.

Chen B, Kubelik AR, Mohr S, Breitenberger CA. **1996.** Cloning and characterization of the neurospora crassa cyt-5 Gene - A nuclear-coded mitochondrial, RNA polymerase with a polyglutamine repeat. *Journal of Biological Chemistry* **271:** 6537-6544.

Dombrowski S, Hoffmann M, Guha C, Binder S. **1999.** Continuous primary sequence requirements in the 18-nucleotide promoter of dicot plant mitochondria. *Journal of Biological Chemistry* **274:** 10094-10099.

Duff RJ, Nickrent DL. **1999.** Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. *American Journal of Botany* **86:** 372-386.

Emanuel C, von Groll U, Muller M, Borner T, Weihe A. **2006.** Development- and tissue-specific expression of the RpoT gene family of Arabidopsis encoding mitochondrial and plastid RNA polymerases. *Planta* **223:** 998-1009.

Emanuel C, Weihe A, Graner A, Hess WR, Borner T. **2004.** Chloroplast development affects expression of phage-type RNA polymerases in barley leaves. *Plant Journal* **38:** 460-472.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. **2000.** Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300:** 1005-1016.

Filθe J, Forterre P. **2005.** Viral proteins functioning in organelles: a cryptic origin? *Trends in Microbiology* **13:** 510-513.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. **1999.** Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151:** 1531-1545.

Gordon K, Futterer J, Hohn T. **1992.** Efficient Initiation of Translation at Non-Aug Triplets in Plant-Cells. *Plant Journal* **2:** 809-813.

Gray MW. **1989.** The Evolutionary Origins of Organelles. *Trends in Genetics* **5:** 294-299.

Gray MW. **1992.** The endosymbiont hypothesis revisited. *Int.Rev.Cytol.* **141:** 233-357.

Gray MW. **1993.** Origin and evolution of organelle genomes. *Curr.Opin.Genet.Dev.* **3:** 884-890.

Greenleaf AL, Kelly JL, Lehman IR. **1986.** Yeast Rpo41 Gene-Product Is Required for Transcription and Maintenance of the Mitochondrial Genome. *Proceedings of the National Academy of Sciences of the United States of America* **83:** 3391-3394.

Hedtke B, Borner T, Weihe A. **1997.** Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. *Science* **277:** 809-811.

Hedtke B, Borner T, Weihe A. **2000.** One RNA polymerase serving two genomes. *Embo Reports* **1:** 435-440.

Hedtke B, Legen J, Weihe A, Herrmann RG, Borner T. **2002.** Six active phage-type RNA polymerase genes in Nicotiana tabacum. *Plant Journal* **30:** 625-637.

Hedtke B, Meixner M, Gillandt S, Richter E, Borner T, Weihe A. **1999.** Green fluorescent protein as a marker to investigate targeting of organellar RNA polymerases of higher plants in vivo. *Plant Journal* **17:** 557-561.

Hess WR, Borner T. **1999.** *Organellar RNA polymerases of higher plants*. SAN DIEGO: ACADEMIC PRESS INC.

Hess WR, Prombona A, Fieder B, Subramanian AR, Borner T. **1993.** Chloroplast Rps15 and the Rpob/C1/C2 Gene-Cluster Are Strongly Transcribed in Ribosome-Deficient Plastids - Evidence for A Functioning Non-Chloroplast-Encoded Rna-Polymerase. *Embo Journal* **12:** 563-571.

Huelsenbeck JP, Ronquist F. **2001.** MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17:** 754-755.

Igloi GL, Kossel H. **1992.** The Transcriptional Apparatus of Chloroplasts. *Critical Reviews in Plant Sciences* **10:** 525-558.

Ikeda TM, Gray MW. **1999.** Identification and characterization of T3/T7 bacteriophage-like RNA polymerase sequences in wheat. *Plant Molecular Biology* **40:** 567-578.

Isono K, Shimizu M, Yoshimoto K, Niwa Y, Satoh K, Yokota A, Kobayashi H. **1997.** Leaf-specifically expressed genes for polypeptides destined for chloroplasts with domains of sigma(70) factors of bacterial RNA polymerases in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America* **94:** 14948-14953.

Joshi CP, Zhou H, Huang XQ, Chiang VL. **1997.** Context sequences of translation initiation codon in plants. *Plant Molecular Biology* **35:** 993-1001.

Kabeya Y, Hashimoto K, Sato N. **2002.** Identification and characterization of two phage-type RNA polymerase cDNAs in the moss Physcomitrella patens: Implication of recent evolution of nuclear-encoded RNA polymerase of plastids in plants. *Plant and Cell Physiology* **43:** 245-255.

Kabeya Y, Sato N. **2005.** Unique translation initiation at the second AUG codon determines mitochondrial localization of the phage-type RNA polymerases in the moss Physcomitrella patens. *Plant Physiology* **138:** 369-382.

Kapoor S, Sugiura M. **1999.** Identification of two essential sequence elements in the nonconsensus type II PatpB-290 plastid promoter by using plastid transcription extracts from cultured tobacco BY-2 cells. *Plant Cell* **11:** 1799-1810.

Kapoor S, Suzuki JY, Sugiura M. **1997.** Identification and functional significance of a new class of non-consensus-type plastid promoters. *Plant Journal* **11:** 327-337.

Kelly JL, Greenleaf AL, Lehman IR. **1986.** Isolation of the Nuclear Gene Encoding A Subunit of the Yeast Mitochondrial Rna-Polymerase. *Journal of Biological Chemistry* **261:** 348-351.

Kestermann M, Neukirchen S, Kloppstech K, Link G. **1998.** Sequence and expression characteristics of a nuclear-encoded chloroplast sigma factor from mustard (Sinapis alba). *Nucleic Acids Res.* **26:** 2747-2753.

Kim M, Mullet JE. **1995.** Identification of a sequence-specific DNA binding factor required for transcription of the barley chloroplast blue light-responsive psbD-psbC promoter. *Plant Cell* **7:** 1445-1457.

Kobayashi Y, Dokiya Y, Kumazawa Y, Sugita M. **2002.** Non-AUG translation initiation of mRNA encoding plastid-targeted phage-type RNA polymerase in Nicotiana sylvestris. *Biochemical and Biophysical Research Communications* **299:** 57-61.

Kobayashi Y, Dokiya Y, Sugita M. **2001a.** Dual targeting of phage-type RNA polymerase to both mitochondria and plastids is due to alternative translation initiation in single transcripts. *Biochemical and Biophysical Research Communications* **289:** 1106-1113.

Kobayashi Y, Dokiya Y, Sugiura M, Niwa Y, Sugita M. **2001b.** Genomic organization and organ-specific expression of a nuclear gene encoding phage-type RNA polymerase in Nicotiana sylvestris. *Gene* **279:** 33-40.

Kohler RH, Zipfel WR, Webb WW, Hanson MR. **1997.** The green fluorescent protein as a marker to visualize plant mitochondria in vivo. *Plant Journal* **11:** 613-621.

Kuhn K, Bohne AV, Liere K, Weihe A, Borner T. **2007.** Arabidopsis phage-type RNA polymerases: Accurate in vitro transcription of organellar genes. *Plant Cell* **19:** 959-971.

Kuhn K, Richter U, Meyer EH, Delannoy E, de Longevialle AF, O'Toole N, Borner T, Millar AH, Small ID, Whelan J. **2009.** Phage-Type RNA Polymerase RPOTmp Performs Gene-Specific Transcription in Mitochondria of Arabidopsis thaliana. *Plant Cell* **21:** 2762-2779.

Kuhn K, Weihe A, Borner T. **2005.** Multiple promoters are a common feature of mitochondrial genes in Arabidopsis. *Nucleic Acids Research* **33:** 337-346.

Kusumi K, Komori H, Satoh H, Iba K. **2000.** Characterization of a zebra mutant of rice with increased susceptibility to light stress. *Plant and Cell Physiology* **41:** 158-164.

Kusumi K, Yara A, Mitsui N, Tozawa Y, Iba K. **2004.** Characterization of a rice nuclear-encoded plastid RNA polymerase gene OsRpoTp. *Plant and Cell Physiology* **45:** 1194-1201.

Kusumi K, Mizutani A, Nishimura M, Iba K. **1997.** A virescent gene V1 determines the expression timing of plastid genes for transcription/translation apparatus during early leaf development in rice. *The Plant Journal* **12:** 1241-1250.

Lang BF, Burger G, Okelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW. **1997.** An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* **387:** 493-497.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. **2007.** Clustal W and clustal X version 2.0. *Bioinformatics* **23:** 2947-2948.

Lerbsmache S. **1993.** The 110-Kda Polypeptide of Spinach Plastid Dna-Dependent Rna-Polymerase - Single-Subunit Enzyme Or Catalytic Core of Multimeric Enzyme Complexes. *Proceedings of the National Academy of Sciences of the United States of America* **90:** 5509-5513.

Li JN, Maga JA, Cermakian N, Cedergren R, Feagin JE. **2001.** Identification and characterization of a Plasmodium falciparum RNA polymerase gene with similarity to mitochondrial RNA polymerases. *Molecular and Biochemical Parasitology* **113:** 261-269.

Liere K, Borner T. 2007. Transcription of plastid genes. In:*Regulation of Transcription in Plants.* Oxford: Blackwell, 184-224.

Liere K, Kaden D, Maliga P, Borner T. **2004.** Overexpression of phage-type RNA polymerase RpoTp in tobacco demonstrates its role in chloroplast transcription by recognizing a distinct promoter type. *Nucleic Acids Research* **32:** 1159-1165.

Liere K, Maliga P. **1999.** In vitro characterization of the tobacco rpoB promoter reveals a core sequence motif conserved between phage-type plastid and plant mitochondrial promoters. *Embo Journal* **18:** 249-257.

Lin ZG, Kong HZ, Nei M, Ma H. **2006.** Origins and evolution of the recA/RAD51 gene family: Evidence for ancient gene duplication and endosymbiotic gene transfer. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 10328-10333.

Lupold DS, Caoile AGFS, Stern DB. **1999.** The maize mitochondrial cox2 gene has five promoters in two genomic regions, including a complex promoter consisting of seven overlapping units. *Journal of Biological Chemistry* **274:** 3897-3903.

Lynch M, Conery JS. **2000.** The evolutionary fate and consequences of duplicate genes. *Science* **290:** 1151-1155.

Masters BS, Stohl LL, Clayton DA. **1987.** Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell* **51:** 89-99.

Matsunaga M, Jaehning JA. **2004a.** A mutation in the yeast mitochondrial core RNA polymerase, Rpo41, confers defects in both specificity factor interaction and promoter utilization. *Journal of Biological Chemistry* **279:** 2012-2019.

Matsunaga M, Jaehning JA. **2004b.** Intrinsic promoter recognition by a "core" RNA polymerase. *Journal of Biological Chemistry* **279:** 44239-44242.

Mcallister WT, Raskin CA. **1993.** The Phage Rna-Polymerases Are Related to Dna-Polymerases and Reverse Transcriptases. *Molecular Microbiology* **10:** 1-6.

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K. **1992.** Gene Organization Deduced from the Complete Sequence of Liverwort Marchantia-Polymorpha Mitochondrial-Dna - A Primitive Form of Plant Mitochondrial Genome. *Journal of Molecular Biology* **223:** 1-7.

Oikawa K, Tanaka K, Takahashi H. **1998.** Two types of differentially photo-regulated nuclear genes that encode sigma factors for chloroplast RNA polymerase in the red alga Cyanidium caldarium strain RK-1. *Gene* **210:** 277-285.

Peeters NM, Chapron A, Giritch A, Grandjean O, Lancelin D, Lhomme T, Vivrel A, Small I. **2000.** Duplication and quadruplication of Arabidopsis thaliana cysteinyl- and asparaginyl-tRNA synthetase genes of organellar origin. *Journal of Molecular Evolution* **50:** 413-423.

Phillips TL, DiMichele WA. **1992.** Comparative ecology and life-history biology of arborscent lycopsids in late Carboniferous swamps of Euramerica. *Annals Missouri Botanical Garden* **79:** 560-588.

Qiu YL, Li LB, Wang B, Chen ZD, Dombrovska O, Lee J, Kent L, Li RQ, Jobson RW, Hendry TA, Taylor DW, Testa CM, Ambros M. **2007.** A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. *International Journal of Plant Sciences* **168:** 691-708.

Rapp WD, Stern DB. **1992.** A Conserved 11 Nucleotide-Sequence Contains An Essential Promoter Element of the Maize Mitochondrial Atp1 Gene. *Embo Journal* **11:** 1065-1073.

Raubeson LA, Jansen RK. **1992.** Chloroplast Dna Evidence on the Ancient Evolutionary Split in Vascular Land Plants. *Science* **255:** 1697-1699.

Reichel C, Mathur J, Eckes P, Langenkemper K, Koncz C, Schell J, Reiss B, Maas C. **1996.** Enhanced green fluorescence by the expression of an Aequorea victoria green fluorescent protein mutant in mono- and dicotyledonous plant cells. *Proceedings of the National Academy of Sciences of the United States of America* **93:** 5888-5893.

Renzette N, Gumlaw N, Nordman JT, Krieger M, Yeh SP, Long E, Centore R, Boonsombat R, Sandler SJ. **2005.** Localization of RecA in Escherichia coli K-12 using RecA-GFP. *Molecular Microbiology* **57:** 1074-1085.

Richter U, Kiessling J, Hedtke B, Decker E, Reski R, Borner T, Weihe A. **2002.** Two RpoT genes of Physcomitrella patens encode phage-type RNA polymerases with dual targeting to mitochondria and plastids. *Gene* **290:** 95-105.

Riechmann JL, Ito T, Meyerowitz EM. **1999.** Non-AUG initiation of AGAMOUS mRNA translation in Arabidopsis thaliana. *Mol.Cell Biol.* **19:** 8505-8512.

Ronquist F, Huelsenbeck JP. **2003.** MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572-1574.

Sambrook J, Russell DW. **2001.** *Molecular Cloning: a Laboratory Manual.* Cold Spring Harbour Laboratory Press.

Schmidt HA, Strimmer K, Vingron M, von Haeseler A. **2002.** TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18:** 502-504.

Shiina T, Tsunoyama Y, Nakahira Y, Khan MS. **2005.** *Plastid RNA polymerases, promoters, and transcription regulators in higher plants*. SAN DIEGO: ELSEVIER ACADEMIC PRESS INC.

Shutt TE, Gray MW. **2006.** Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends in Genetics* **22:** 90-95.

Silhavy D, Maliga P. **1998.** Mapping of promoters for the nucleus encoded plastid RNA polymerase (NEP) in the iojap maize mutant. *Current Genetics* **33:** 340-344.

Skinner. **2006.** The PLANTS Database, Version 3.5. *National Plant Data Center USDA NRCS Baton Rouge LA.*

Small I, Peeters N, Legeai F, Lurin C. **2004.** Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **4:** 1581-1590.

Sousa R, Chung YJ, Rose JP, Wang BC. **1993.** Crystal-Structure of Bacteriophage-T7 Rna-Polymerase at 3.3-Angstrom Resolution. *Nature* **364:** 593-599.

Swiatecka-Hagenbruch M, Emanuel C, Hedtke B, Liere K, Borner T. **2008.** Impaired function of the phage-type RNA polymerase RpoTp in transcription of chloroplast genes is compensated by a second phage-type RNA polymerase. *Nucleic Acids Research* **36:** 785-792.

Swiatecka-Hagenbruch M, Liere K, Borner T. **2007.** High diversity of plastidial promoters in Arabidopsis thaliana. *Molecular Genetics and Genomics* **277:** 725-734.

Tan XY, Liu XL, Wang W, Jia DJ, Chen LQ, Zhang XQ, Ye D. **2010.** Mutations in the Arabidopsis Nuclear-Encoded Mitochondrial Phage-Type RNA Polymerase Gene RPOTm Led to Defects in Pollen Tube Growth, Female Gametogenesis and Embryogenesis. *Plant and Cell Physiology* **51:** 635-649.

Tanaka K, Oikawa K, Ohta N, Kuroiwa H, Kuroiwa T, Takahashi H. **1996.** Nuclear encoding of a chloroplast RNA polymerase sigma subunit in a red alga. *Science* **272:** 1932-1935.

Tanaka K, Tozawa Y, Mochizuki N, Shinozaki K, Nagatani A, Wakasa K, Takahashi H. **1997.** Characterization of three cDNA species encoding plastid RNA polymerase sigma factors in Arabidopsis thaliana: Evidence for the sigma factor heterogeneity in higher plant plastids. *Febs Letters* **413:** 309-313.

Taylor JS, Raes J. **2004.** Duplication and divergence: The evolution of new genes and old ideas. *Annual Review of Genetics* **38:** 615-643.

Tiranti V, Savoia A, Forti F, D'Apolito MF, Centra M, Rocchi M, Zeviani M. **1997.** Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPOL) by cyberscreening of the Expressed Sequence Tags database. *Hum.Mol.Genet.* **6:** 615-625.

Trifa Y, Privat I, Gagnon J, Baeza L, Lerbs-Mache S. **1998.** The nuclear RPL4 gene encodes a chloroplast protein that co-purifies with the T7-like transcription complex as well as plastid ribosomes. *Journal of Biological Chemistry* **273:** 3980-3985.

Unseld M, Marienfeld JR, Brandt P, Brennicke A. **1997.** The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. *Nature Genetics* **15:** 57-61.

Wagner A. **1994.** Evolution of Gene Networks by Gene Duplications - A Mathematical-Model and Its Implications on Genome Organization. *Proceedings of the National Academy of Sciences of the United States of America* **91:** 4387-4391.

Wang GF, Kong HZ, Sun YJ, Zhang XH, Zhang W, Altman N, dePamphilis CW, Ma H. **2004a.** Genome-wide analysis of the cyclin family in Arabidopsis and comparative phylogenetic analysis of plant cyclin-like proteins. *Plant Physiology* **135:** 1084-1099.

Wang JPZ, Lindsay BG, Leebens-Mack J, Cui LY, Wall K, Miller WC, dePamphilis CW. **2004b.** EST clustering error evaluation and correction. *Bioinformatics* **20:** 2973-2984.

Wang WM, Tanurdzic M, Luo MZ, Sisneros N, Kim HR, Weng JK, Kudrna D, Mueller C, Arumuganathan K, Carlson J, Chapple C, de Pamphilis C, Mandoli D, Tomkins J, Wing RA, Banks JA. **2005.** Construction of a bacterial artificial chromosome library from the spikemoss Selaginella moellendorffii: a new resource for plant comparative genomics. *Bmc Plant Biology* **5**.

Weihe A. **2004.** In: Daniell H, Chase CD, eds. *The transcription of plant organelle genomes*. Berlin: Springer, 213-237.

Weihe A, Borner T. **1999.** Transcription and the architecture of promoters in chloroplasts. *Trends in Plant Science* **4:** 169-170.

Weihe A, Hedtke B, Borner T. **1997.** Cloning and characterization of a cDNA encoding a bacteriophage-type RNA polymerase from the higher plant Chenopodium album. *Nucleic Acids Research* **25:** 2319-2325.

Weng JK, Tanurdzic M, Chapple C. **2005.** Functional analysis and comparative genomics of expressed sequence tags from the lycophyte Selaginella moellendorffii. *Bmc Genomics* **6**.

Williams S. **1937.** Correlation phenomena and hormones in Selaginella. *Nature* **139:** 966.

Xu GX, Ma H, Nei M, Kong HZ. **2009.** Evolution of F-box genes in plants: Different modes of sequence divergence and their relationships with functional diversification. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 835-840.

Yoo SD, Cho YH, Sheen J. **2007.** Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat.Protoc.* **2:** 1565-1572.

Young DA, Allen RL, Harvey AJ, Lonsdale DM. **1998.** Characterization of a gene encoding a single-subunit bacteriophage-type RNA polymerase from maize which is alternatively spliced. *Molecular and General Genetics* **260:** 30-37.

Zhang JZ. **2003.** Evolution by gene duplication: an update. *Trends in Ecology & Evolution* **18:** 292-298.

Zhang JZ, Zhang YP, Rosenberg HF. **2002.** Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature Genetics* **30:** 411-415.

Zhou X, Ma H. **2008.** Evolutionary history of histone demethylase families: distinct evolutionary patterns suggest functional divergence. *Bmc Evolutionary Biology* **8**.

6 Appendix

6.1 The genomic DNA sequence of *RpoT* gene from *Selaginella moellendorffii*

6.2 The genomic DNA sequence of *NaRpoTm1*

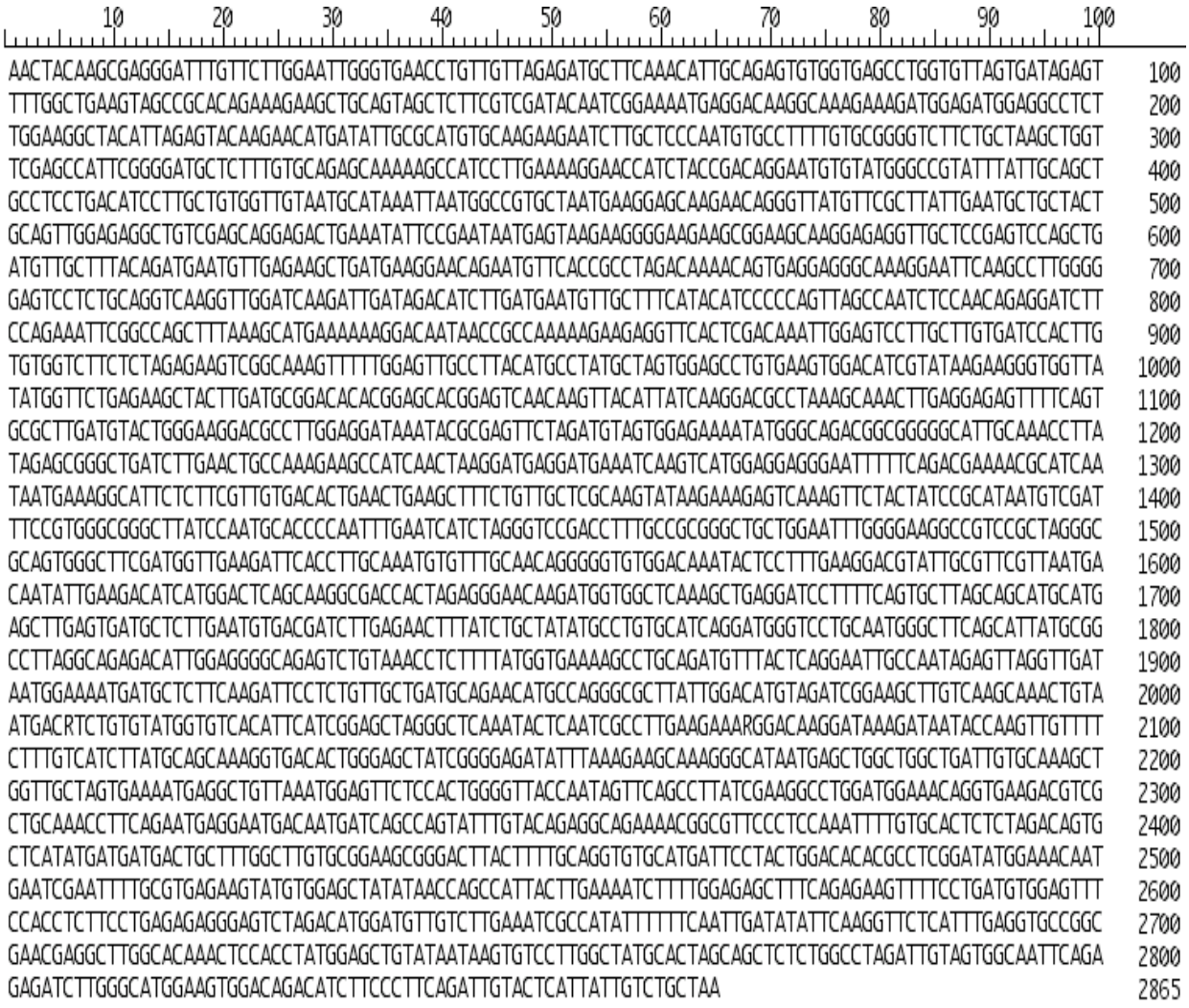
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150

ATATATATATATATATATATATTGCTGCACCCCCCACCCCCGCACCCAAGTTTTTTGAAAGTGCCGCACCCCCACACCCGCACCCATGTGACATAGCTATGTATCATGAATTTACTGTTCAAAAGGTGCTACTAGAAATTAGTGTACAGA 15750
ACCTCAAACAGGAACATAGATGGGTTGGACTGGAGTCTGTCCTAGGCTGATCAGGTTCTAATAAAGCTGCTCATGCTGTTAGTTCTTCGTTGGTATGCTTTGTTATTTGAATTTAGTTTCCATATCTATAAGGGAAACTCGAACCTGACC 15900
AAATTTTCTTCATCCTGACTTTTTTTTTCTATATGCTTCAGTTTCCATTGGCAGGTGTGCATTGGTCTGAATAGGACTTGGATTTACTCGGGCTTACGAGCTAACCTTATCTGATGTAGATAGTCTATTTAGGTGCCTTCSTGTAGGATC 16050
TGGCTTTGTTTCTGATCTGACCAACTACATTCTTACTTTTATGCGATTGTTTAGTTTTGGAAACAAAAAGGTGTCTTTAGTTTTGTGAGAGAAAGAGACATGGGTGCATAAATTCTATGTTATTGTTATTGTAMATATGGATCTTTCCT 16200
AAACTTATGAACCAAAATATTTCAATTTTGTTGATAATCTTGAATTGATGAATATATGCATTGGTAATATGTGCTATCATACTATTCTGGGAATCAGACCTAATTGTAGGCAAARCATATCTGTATGCATTCCCAACCTTTTACATCCTG 16350
CCAAAAAGCTAATTTATCTTTCATCTGATTGGCTTTTCTGAGATTGTCTTATTGTGGATTTATTTAGGATGGCTCTTGTAATGGCTTGCAACACTATGCAGCCCTGGGAAGAGACAAGGTAATGCATAATTATCCTACATAACCATTGAA 16500
ATGCACCTTTCTGAGGTCATTGATGTATTCTTCTTCAAGGTCTCTCATGTATCTGGTGGAATGTGTTCAATTTTGTTCTTCAATTTCATTGCTTGTGTAATTAGAGGCACATTGCATTGATGTTTTGCAACTTATTTGATATCCCTGTTA 16650
TATTTAGTTTTTTTGAGGTTCCTGCATAGAGAAAGAAGACACGAACAAGGTTTGGTCTGAACTCTCACGATTCCCAATATGGGAAAATCATGTCTCATTGCGACTATTTATATTCACTTTACACTAAAATACATCTTTATTAGGAACACA 16800
ATATATAATAAAATACAATAGGGAACTAAGAGACATGAATAACGACGTGAACAGTAACATCATGAACAATGACGTGAGTGATATAGATCTGGATCATGATCCAGATCGGACCACCTCAAACCTCATGCTCTAATAGCCTTTCTCAAGTTG 16950
TACATGATTAACAACCATGGACCACCTGCTTCCTAAATACCAAAACCTGTTGTATTTTATTACTGATTTCAAACTTAGCCCTCGATCGATATCAGGAATCAACACCATCCATCGGAACATTGAAAATCAACAACCAATATCGAGACACTA 17100
TGAAAATCAACAACCATAATGCCACTTTCTTGGTATACCAAAACAACAGTTGTATTTACACACCACCCACAGGCAAAGTGACAGACTCTTCCTAAAACCGTAGTCAGGAGAAAAACCTCAATATCGGGTTCAGGGGCAACAACAACGCAC 17250
ATGAAATTTTAGTATCCTCACAGTTCTCTAAAACTGCAATGCTCATAATTCTCATACTATAGAAGGGACATCAAGGGAACGTAGATTATATTATTTTTTAATTTCTTTTATCTTGTTATTAATTCATCTTCTTTAAATTGTAATTCACCA 17400
TGTAGAATTAATAACACAAGCTCGCAAATCAACATGCACAGCTCTTCATAGGGACACAACCCATGACAACATAGAATAGTGCAATAAAAACAATAATATTACAATTGATTTTTGTAATTCATAGAACAACTAAAACAAAGAGCAATACAT 17550
GCTAAAGGCTCGCACAACCTACGATGACTCCTTTTGATCATCAATGCATTCATATGCAAAAATCGTGAATACATGCTAATATGCATTCATTCCTTTTAAATGCTTCTCAATTCACATTCATTTATGCCAATTAAAAGATGTGCAAGAAAC 17700
AATAATATTCATATGCATGTGCATAAGTTTACACAACTATACCAAGAGCATGCTTTCATTCATGAATCACTTGATTTAAATCATGCGAACACTCATGCAATTATTTACATGCAACAACATATCAATTTTATGTATGCTCATGCATAAAGA 17850
AACAACCTTATATTAAAACTTTAACTCAGCATATATTATCTAGTTTAAGCATGTGAACATTCATGTGATTACTTACATGCAAAACACATCGATTTCAAGAATGCACGTGCAGATAAGAACAATTGTGTATTAACACTTCGATTAAACATG 18000
ATATACTAATGTGCAAGCATGCACATTTTTATTCTTTTACATGCATACAACCAGCCCAATTTGAAATAACAAATTTGGAAATTTTCATGCCAAACTGAATATGCCCTTTTCAAATTCAAATAAGCATATCACATGAGCATCAACTGAAAA 18150
TTTACATGTCTCATCTGAACCTCCACTTATGCATAAACTGAACATGCCAATAATATAAAAAATATATTCTTTTTTTTTGTGTTTTTTTTGCACCTAATGCTGCCCCAGCCGAACAACAGGTCGGCTAGGCCTGTTCCGTGATATAAACC 18300
AGTCCAGCCGAGTCTGACACTGTGTCGAGTACCAAATCATGTTGGACTTCTTGTTCAAGCATGTTTTCCATGGCAGAAGTGTTAAAATATAAAAAATATATATATATAATAATATATTTATATTATTATTATTATATATATATTTTTGTAAA 18450
TATAATTTTTATTTTTATTTATTTATTATCTAGTTCTCATAGAACTAGATAATAACTATTTGTCAAGTTCCAATTGGAATTTGACCATATGCATAGTCGTTGGACCTAGTCCAACGGCTACAAGGAGAGCTCCTTAAGAGCTCTCCTCTA 18600
ACCCCTTCTCACTTGTATAAATAGGGGTGCATCCCCTCATCTAAAATCATTCTCAATGCTTGTTAAGCATTTTCTTGTTATTAATAAAGTTTTTTTTATGTTATTCCATTTATACCCTTTTGTGGGTCTTGTCTAAAATTGTCTTTTAGA 18750
CCCATCATAGATCGTAGAGTAAATATTCAGATTTACTCTTCCACATGGATTTCTACTAAAACTCTGCATGGAAATGAATCCATGCTATTTTTCTCATTCGAAACAAGAGAAAGACCCATACTTGATCGTAGATCGATAAAAAATAAGAAA 18900
ATAATGATAAAAACGATGGATGATTGGAGAAAAAGCTCCCTTCTTCTTCTTCCTCTTCCTCTTCCTCTCATTTCCTCTCCTTTTCTTTCTTTCTGTCTTTTTCCTTCTCTTTCTCTTTTTTTTAATCTATTTATAGAACTCTAGGGGGGG 19050
GCACCGCCCCCTTCTCCAGCCCCTTCTAGATGGGCTGGAGAAGACAGTGCATCTCCACCTAGAACTCTCTTCTCTCTTGTGAGTTTCGAAAATTTTAAAATCCTTCAAGTTGACTTTTCAAATTTTTATTTTAATAGGCTTTCTAAAACC 19200
TTCTTAGAGGCTTTCGTAAATGTTTAAGGACTCTTTTAAGAGTCCTAAACCAAAGCCAAATTAAAATTTCGAAATTAAATAAAATTAATTATTAAATATTAATCAATTAGTAAATACTTAGAAGCTTCCAAAAGTTATTTTAATAACATA 19350
GAAGACTTCTTAACACTAAATAACTCCAATTTAAAGCTATCTAATGAATTTAAAGTTGGAAATTTGATGCGCATTGATTTAGATCTGGATCCAGATCTAAGATCCAAATCCATCTCAAAAATCTAAATGGGTCGGGCTGGATCAAATCCA 19500
GATTAATATGGGTCTTGGATAAAGTTCAGATTCATATTAATTTGGATCAGATCCAGTCTGATCCAAATCCGGATCCGAATCGACAAACTACTCATTTTTCGAAAAAAATACTATTTTTGGGCCAACCGTATCTTGAATCAAGGAAAATAAA 19650
ATATCGTTACGCTCAGAATTTTGAGACGCATCCAACGATATATAACTCGAATTTCGCTTCCAAAATTCGCCCAAAAATAATCTAAATTTTGCTCGAATTTAAAGGTCGAAAATAGAACTTTTTCTCACGTTTTGAGTGGCTCGTTTAGGC 19800
CTTGACAATGAGACCTGGGCCAAGCCCAGACTCCAAAGCATTGGCCTAAAATCGAGAATCTACCGATTTTTCTTCTATACTTAGCCAAATTTACAAAATCTCAAATTTAACTAAGTGTAAAAATACACGATATTATAATTTTCGCACTAC 19950
AATTCTCTTGCCTTTACTAAAATCCGAGCTGCTTGTTGCTTCATGACTGTCCCCTTTATTTTAGCATTCAACATGTTTAGAGTTCTAGTTTTGCCTTTTGTCCTTGCAATACATTGACATCAGAGAATGAAGAAGTGTTATTTATGACTC 20100
TCTCTCTCTCTCTCTCTCTCTCTATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATTGCTCATGTGCACATACATCTTTAGGATTCATTTACTAGTTAGATTGAAAAATAAATAA 20250
TTATTTTGTTAATTAAAGTTGTAACATTATTGTGGTGCTTAAACTTTTGTTTTGTGAGACTGTATCTGATATATTGCATACCACATAATTGTTATATTCAGCTGGGAGCAGCTGCAGTGAACCTGGTTGCTGGGGAGAAGCCGGCGGATG 20400
TTTACTCCGGAATTGCTGAAAGGTGTGCAAAATATCTATTTGTGTGGTCTTTGTTCTGTTCTTTTTTTGACTTGATTTGATTTTTTTTAATTAGTCTATGTTTGGTTTTGGTGAAGATGAATTCCAATGAAAAATGAGTTCCTTGGAACT 20550
TCCCATTCAAATTCTTCAGTGTGTAACTAAAATGGGCACCTACATGTGCTGTTTGATTTCTGCTAACTTGATGGCAAATGGGCTTCTCAGAGCTGGTACAAGATCATGTTGATGAAAGTATTTTGTGGAGTCCAAAGACAATCTTCATGT 20700
CATCATATCACCAAATTGGTTCATTTTATTTATTTTTAGATAGGCTACCCTTGTGGTTCACTTTTATCCATTATTTGTTTTGACTATATTTACGCAAATGGGTTCTGTTCATTAAGCCTAATTCACTTGTGTTCCTGTAAATTTATCCTT 20850
CGTTGTTCTTGTATGAATGATTGCATACAGTTTCTAGTTTGTGTATATTCAGATATTGACCTCAATCATGATTTAGTTGTGTAGTTCACATGAAAGACTACTGCTTATAATATGCCAGATTTGTCTCTACAACTGTAAATGCCACATGCC 21000
TTTAGATTCCTTCTTACACATATAACTGGAGCCTGTCCAAATGAACTGCCCTTGTTTGAAGATTACAATTATCAATATAATTCTTTTTCTGGTGCTTCTTTTGCATTTTGGGGATGTGAAGAATGATATGGTTCTCATATTATTGTTGGA 21150
GGGTTTTAAAAGCCAATTTTGCCTCTCTCTCTCTCTCTCTCTCTTTTTCCTTCCACAAACATTGTCTGGATTCCTGATTACATGATGATATGGGTCAGCTCATCTTTATCTTGGTTGGTCATCCATAATTTTTTCACTGTTTGACAGTCA 21300
GATTTTTGTGAAGCAGATAGTTTTTACCTGTACTTGAGAGCATAGACGCCTGACACTCTTTTTGTAACATTTGGTTTCCACTGGACTGCCTGTCTCATGATTAATTGAATATCTGTCATACTCAGATTTCCTGGTCAACTCTATCACATT 21450
GACTGGGTGCACCCTTGTACATAGATAAATGGGATAAGTGGGAACTTGCATTAGGTGTTGCAGTAGAAAATTCCCTGCTTAGCCCATTTATCTAGATAGTGTGCATCCATACCCCCTTGTAGGAAAAAGAAAGAAAAAATAATGTCAAAT 21600
TGGAGTATATAAAATAGGAAGATCTAAAAGTGAAGAAGAAAATGAATAAAGATACACCTACCAAATTTAAAAAGGATCAAAAACTAAAATATTGGAAATAACAGTGCATTTAAACACTTTCTTTTGAGCATCAGAGCACTTAAATATTGA 21750
TTCTGAACACTAAAAAGAATGTGAGTTTTATTCCTGTTTCTCATGACACAAATATTTCAATCTGGTGATATTTCCTCGAAATTTAAAAGCCAAATTCTTGGTATATTTGTGGAACCAATCATTCTTTTCTACCAAGGAAATACACCTTTT 21900
TTTATGTGGCAGTCTATGTCATATAGGTGCGGGTGCAAGTGCAACGATGCGGCACTTTAAAAAAACTTGGGTGAGGGGATGCGGGATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 22050
TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAACTTGGGTCGGGGGCTTGAGGTACCCCTATCCATCTTGAGGTACTCTCGAGCCCCTCCCTTTTTTTAATGAACCAAAATGGACT 22200
CGCTCAAAGTTGAGCCGAGTCCATTTTGGACAAAAACAACCCTATGGTTGACCTCACCCAGTGTCACATTGACCGCACCCGCGTTGCATCGCGTCGTGGCGACTTGAGTGCGGCGGTCTCTATGGCGAGTCCCTGTGACATAAGTGGCAG 22350
TTCTAAGGAAATATATCTTCTTAATTCTAGATCAATATTTCATGTCAAAATTGCATCTTAGCATTGCAGATGCATGGTTAAACTAGTCAGAAATTCAGTCTTGACAATATCTCAATGAGGGTTTTTCTTTAGAGACCTGATTTTTATGTT 22500
CACGTGATTTCACCTGTTGCGTCGTAAACAATTTTCCATAGTGCTTTCTGCTCCCTTCCCGCTTCCTGAATCCTTCTAAAAATTGAATAATGCTGTGATTCCAATTCCATATGCTTCTGCTGATTCTGCTTTAGTAACGGTGTAAATAGC 22650
ACTAAAAATTCATAACTGAAACTCAAAAAATGAATTTCTCAGAAAATAGTGATAATGTGTTATATTTTCATTTTTTCTCTATTGTTTTTTTAAATGTAATGCTTCCGTACTATCACCGGTACAATTGATTTGGATTAAAGTCAGATTTGG 22800
GTTAAAGTCAGCACTCGTGAAATTTGAAGAGTTTATTCTGGAAATAGTTCATAGATTGAGCACAATATTTAGCCAACTGTTTTTATTTTTTCCATTTTCTTAAGTACTTGCCTAAATGAAAAATGGATTAATCCAGAATAATTAGTGTGG 22950
TTAAAGGAGCAGCAAACAATGGTTCTTGTAGGCAATGCAACATGTAATTTTTTCTCTCCTTTTGCATGCCATAGAATCTGTATGCTTTAAAGTGTGATGATCCTAATTTATACTTGGGTCAAGATAGCTCAGGAATGTTTTTGTTGATAA 23100
TTTCCCCAAGTGAGCATTCATGAGTCATGATGCTCATCATCTCAAAATATTTACCAGCAAAATTCCTTTTTATTGTAGATCTCTTACTAAGCTTATTACCTAATAATAAACTGTCATTTTCAGGTAGAGTTTTTGACATCATGAGAAGGG 23250
ATTCCCAAAAAGATCCTACAAATAACCCAAGTGCTGCACGTGCAAGGCTTCTGATTGATCAGGTTGGTTTCTATTCTACACACTGTGCTACTATTTAGTCTTCAGGCTCTGTTTGGGAAGTGGCCGTTGGAAAAATCAAGAGACAAATGG 23400
TTTCTGGAAATGAAGCTTTCTAATCCAACATCGTGTTTGAGAAAGACGAATTGTTAAAAAAGTGGAAAAAAGAAGGTATTTCACATCAACCAACTGATCTCTTTTTCCCAGTTTGGTTTCCTTCTTGGGAAGGAGATAATTATCAATTCT 23550
TTCACCAAGAAGAATCTGATGGATTTGGGGAACTTTTGACCACTAGAAGTCCTATTTTCCTTTCTCATGACATGCAAAATATGCCCTGTTCGTCTTTTTTGCCTCACAAATGAAGATGAGAAAGCAGCATAACCATGGTCCATGGGTAGT 23700
CTTTGAATTGTAATTGGATTTATTTCAATTCTTTTGCTAGATACAACCAAGCATGTAAACAAGAAATTGACGGTCAAACAAGATAGAGAAACAAAGAGAATTGGTAGCCTGATTCCTTGGAAGAATTATAAGGTTTCCAAACAAGATCAA 23850
TTTCCCATCAGTTCTTTCCAAATCCATTTCTTTTCATTTCCCTCCAATTGGTTTCTGCCAAATCCTCTCAACTATCACCTCCCAAACAGATTCTTAAACAAACTATGGACAGTTCCAGGAGCAACAAGTCTGAATTGGGTCACTGTTTTA 24000
CAGAGATTCTATGTTAATTAATTCTTTCATCTATTCTCTCACATCCATCTATATAAATTACTCTTATGAAAGAATGTGAAAGACACAAACAATATATGCTGTTGTGTCTGTTTTAATCTTTTTGGTAAATTGAGTCAGAATGTGGAAGGG 24150
AAATGTAATCAATAATCAAGTGTGACTTAAGAAATGTTTTCCGTTTTAAGAGGTTCATTTCTCCTTTTTATTCTTGAATATGTGCAAGATACTGGAAGTATTTGATGTACTTTTTTGCTATGTAGGTGGACCGTAAATTAGTTAAGCAGA 24300
CAGTTATGACATCAGTCTATGGAGTAACCTATATTGGTGCTCGAGACCAGATAAAAAGGAGGCTAAAGGAGCGGGATGCCATTCCAGATGATCAAGTGTTTGGTGCAGCTTGCTATGCAGCTAAAGTAAGGGCATCAATTAAATTTCTAT 24450
TGACTTAACCTTTATAAGTTTGCTTCACAAAAAAAAAAAAAACAGAAACATTATTTTCAAGTTATACTGTTAAATTTGTCTATCTGCAATGTTCTTAATGCAAGGAAACAATGTGGAATGACGAACAAATTTTTGTATTGAATAGACAAC 24600
CTTGACTGCCTTGGAGGAGATGTTTCAAGCTGCACGTGCCATTATGAGCTGGCTGGGGGATTGTGCAAAGGTTTGTGCATGTTGTCTGTTAATATGATAATTCCTTGAAATGTTCATGATTTTTAATGAACATATTATACCAGAATTCTT 24750
TGAGCTTCTACATGCATCTGCCATCTGAACTCATACTGCTATTCACCAATTTATGGTCATTGTACATCAGTGTTTCAATTTTTGGGTACTTCTAGTCATATATTGTCATCATCAGGATGTTTTTAAGTTTAGAGCGAAAGGCACAATTTA 24900
GAGGGTTGTATTTATGTTGCCTTTGGTTTCTAGATTATCTCACAGGAGCCAGATAACTTGTAGAATGAATATCCTAGGCATTATCCATAGTTTATTTTGAGGGAATTTTTGGGCTTCACCATGACTTAGGAAAGGCTTAACCGCACATAT 25050
ATTTCTCCATAAAAAAACTTCAGTAGATAGATGCGATAGATGTATTGCAAGCTTGCAATTGTGAGTCCCGTTTTGCAGGTTTATGATGCTTTAGTTTTCCTTATCACCGGATTGGGGATCTTCTTTGTGGAAGCCTTTTTAGATATCTTA 25200
TATTCTTTTATGTTTTTTTCTCGTGAATTCATAGGAAGATTGCTACAAGGATCTGAAACAAATAAAGTTTCTCTGAAAGGAGCCCTACCTTCCTAGATGTTCATAAAGTAAGCATGTAGGCAGTGCATGTTTCAGTATATCATGATGTTG 25350
CTGAAGTTTCTCTCTGCATTCGACACCAATTAATTTCCAACTATTGAAATATTATGAAAGTTGGAAGGAATAAAAAATGTTTTGTTTTTGGAATTATGAAATTCATTTAGAATGGGACTTGCAGCATGCTTTGTGTACCCAAATTTTAGT 25500
GGATAATAGCTGGTTCTCATCTGTCTTGATTGCTGACTGGGTTGTATGCATTTATTGATCTACAAACATTAAATATACAAATGGAAGAAGACTGTGATAAGCAAGGGTATGTTAAGAATAATGATGTTGAGGAGTAATGCATGGACCAGT 25650
CCTGCTTGTTTGCTGGGATTCAATAGCTTTCAAGCCTCTGGTGAGGGCTAGATAAGAAATTTTGAATTCTTATCAAACCATACTCTAACCTTCACCTACTTATATTGTTGCTTCTAGGTTGCTGAACATCCATTGTCTGATACAAGCAGA 25800
TGACTGAGGCTGCTCATGAGATGTGGCTATGTTGTAAAATGGTAGTCATGAACTAGGTTACTAGAGTCTCTCCACATGAATCAGACTAGGTCTGATCTTGCTGGCTATGGAGAAAAGTTATGAAAATTGTATAATAAAGTTGGTGAAAGG 25950
AAAACCAGTAACATCATATAATTGGCTATTAATTGCATAATAGATGGGAATATTCATGACCAGACGTATCAAGAAGCATGGAATGGTCATATATGCCTAGGTAAAGATTGGACAAGGACTAGGCTAGAATTGGTATGCTGGGTGGAAACT 26100
AGATTAGCTAGTCCTTCACCTAGTTTTCTTGTAATTAGCTGAACAAAGCAGGCTTGACTAGTTGTTAGGAGGGGGTTCATCCTGACCAGGTTTCAACTAGGGCACTGCATATCTTATAGTTGAATAAGGAGTGTGGGATAATTGGTAGTT 26250
TGAGCAGAAATTACATCTCTCTCTCTCATACGTTTGACAAATATTGGGGCCTTTTCAAAAATTATCATTCGGGATTTCTATTGCAACCCTGAATGTCATTTATAATATGGATGTGGAAGGCCTTGATTTATTCTCCTCGTTTGAGATGCT 26400
TGATTTCCCCTTATTCCAAGATAGCACCATGTTAACAGTTGGGAAGTCAGGTTGCTTTATAGTAGTCAAGATCCTATATTTGAAGTGGTTGTGGAATTTTCATCAAATGGTGGCTATTTTATTATGCTCCGCAGGTGATTGCTTCAGAAA 26550
ATGAGCCTGTGAGGTGGACTACTCCCCTTGGGCTTCCGGTTGTGCAACCTTACAAGAAGAAAGGAAGGTGTCTTGTAAGCTCTCTCTCTCTCTCTCACTTGTGTACAAAAAGTCGGTGAACAGTTTCTTCCAGAAGGTTTGTGCCTTGCA 26700
TCCTTAACCAGTTTCTGAAAGAATGATAAGTTGGAGGCTGAGTGGAATATAAGTTGTGTGAAGGATGTGCGGTTTAGAAGCTCTTTTTCTGCTAAAATGTTGCTCTCTTAGGCACTTTCAAGAACTGTTCACTTTCTATATCACATCGTG 26850
CATATAGTGTGTTTTCTCCATTTCTTAAATTGATGATGAGTTGGTGTTGCAGGTAAAAACTTCTCTTCAGGTTTTGGCGTTAAAAAGAGAAACAGATAAGGTCATCACTAGATTCTTCTTTACCTGAGTGTCTAAATGGAGTTCTTTAAA 27000
TAATATTGATATCATTCTCATCTTTGTTTGTGATACTGACTACATAATTCTGCTACTCAGGTCATGATTAAGAGACAGAGGACTGCTTTCCCCCCAAATTTTGTACATTCCCTCGATGGATCTCACATGATGATGACAGCTGTTGCCTGC 27150
AAAAAGGCAGGATTAAAATTTGCAGGTGCATTTCTTCTTTCATCCATTGACATTTGGGCTGTTTTGAGTTGGACTCTTAGTTTTGTGACTTTGCTTCAGGAGTTCATGATTCATACTGGACACATGCCTGCGATGTRGATGAAATTAACA 27300
GAATACTGAGGGAGAAATTTGTGGAACTCTATGAGCAGCCGATACTGGAAAATGTAAGCACACAAACTTGTCAGAACCAACTTCCATMTATGATTTGTTARAAGTTGTGATAGGCAATAAGGCAAAAAATGAGCCAAAATTCCATTTGAA 27450
GATGGCTGAGTGTGGGTGAGTTAGCTTCCATCATGTCAACAAACCTCTTAGAGAATGGAGAAAAAGTCCAGAGGCCAAGTTTTGAGTTTAAATATATCAAATAGATTAAATGTGTTCTGTGAGTCGATTGTCATCAGCCAAAGTTTTAGA 27600
AGAAATGAATTGTTGATTCATTTGTAGAATAACCTGCATTATRTWGTCATTTTTCTGCAACTGCATTATCTAGTCATTTTTCTACAACCCAAAATTTTGTGATCAGCTTAAAAAATGTGGATTCCTGTCTTTTCAAATAGCAATTGTTCG 27750
CTCTTAAGATTCTTCTTGTTTATCATGCTGTGTTGCTGAACCGTACTCATGCTGATGCAGCTCCTGGAGGGCTTCCAGAAATCATTTCCAAAGTTTAGTTTTCCTCCTCTTCCGGATCGTGGCGATTTTGATCTGAAGGAGGTTTTGCAG 27900
TCAACTTATTTTTTCAACTAATTTATATCCACACCAATTCTTTCTCAAGGGTGGCGAGGGGTACCTGGCTCTGTATGATTGGTCCAATTCTATTTCTGCCGTCACAGTATAAGATGGCAAAATTTCATTCTTTTCTCTAAGGCCTCACCA 28050
ATGGATAGATGAGCCTTGCATAGTCTCTGCCACTGCTGGGCATGCTGAAGATGAATGTAAATAGCACCTAGAGACACCTCTTCAAGAAGGTAAAAGACGTCCGTGACTGCAATTGTACATCTGCTCGTATGTCATTGACCAAAGAGGTAT 28200
GCCTCTGGCATTGTATCTCAATATGGGTTGCCGcACAAACCTTTTKGTGCTTGGTGTCTTCACCTTCCATCTAGTGCCCAATTTTCATTTTGGAGGAARAAGCAAATTGTATGTCAAATAGAGAACATATTACCTACCATAGCTACTGTT 28350
TGGAATGGACAGTGCAATAAAGACGTAATGTTTATGTTCACCGAGTATTAGCATATTCGCTATTGTTGATTTTCCTTAATGCAATCAGTGGCTTGATTTTCAGC 28454

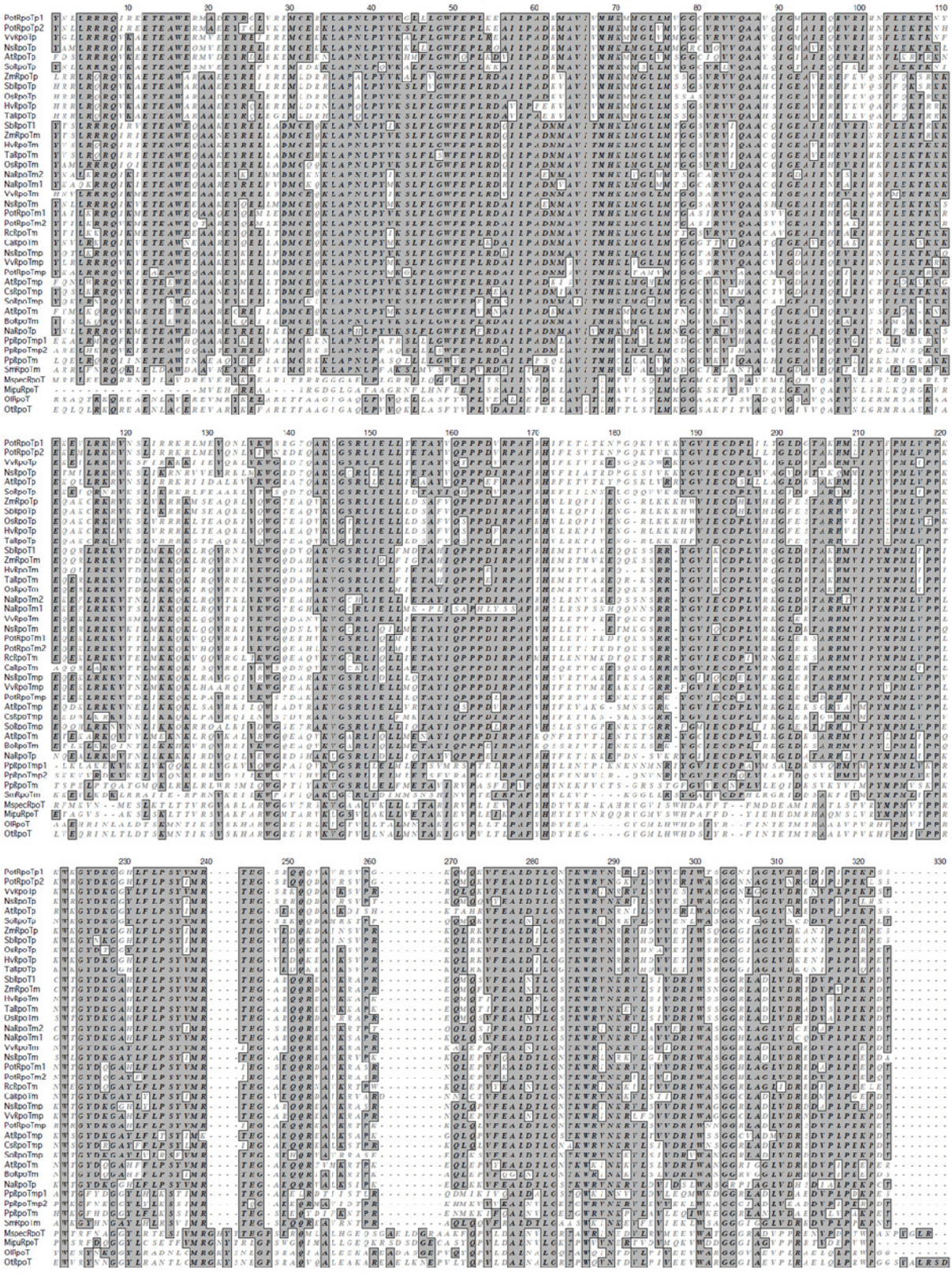
6.3 The genomic DNA sequence of *NaRpoTm2*

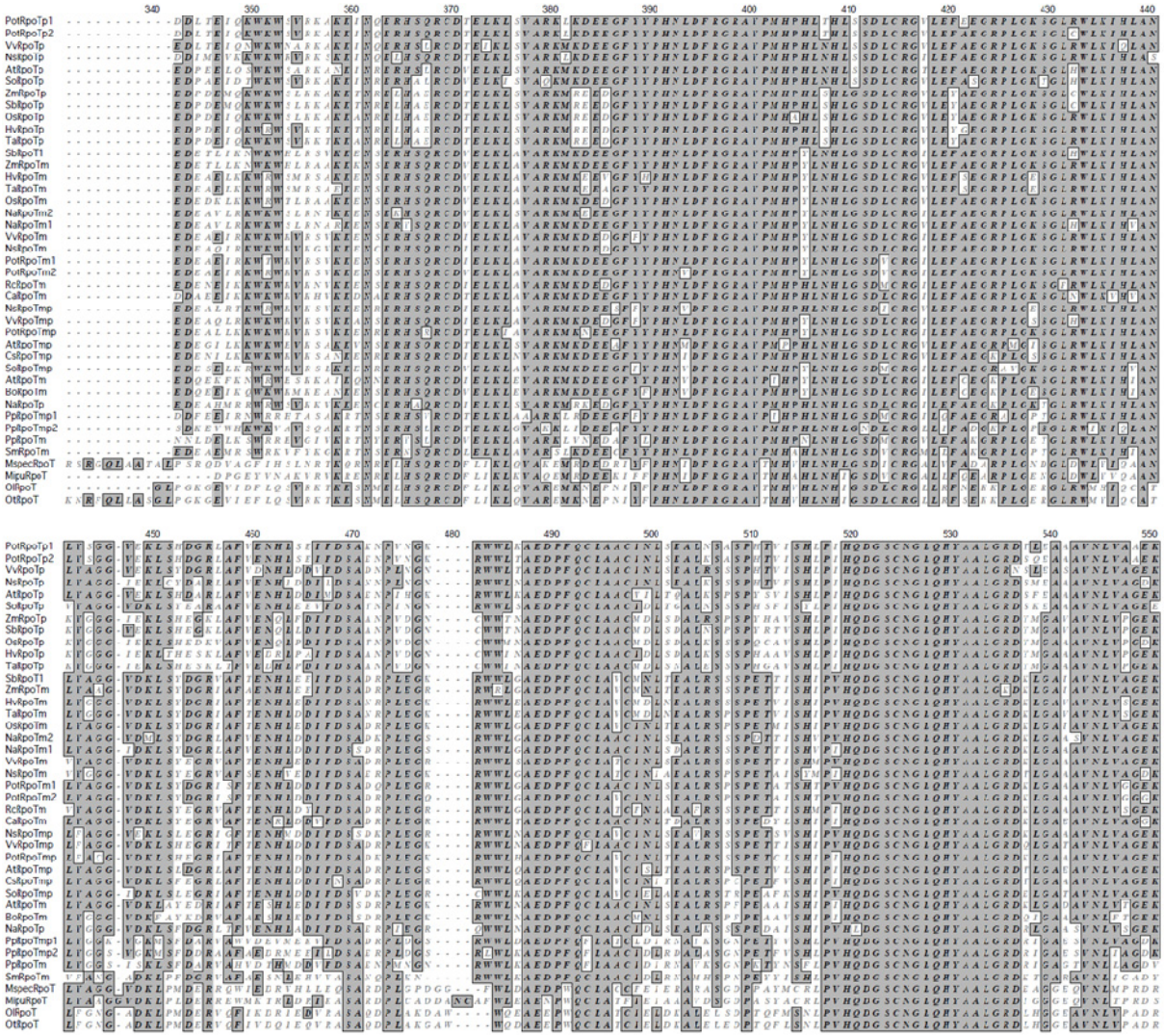
6.4 The genomic DNA sequence of *NaRpoTp*

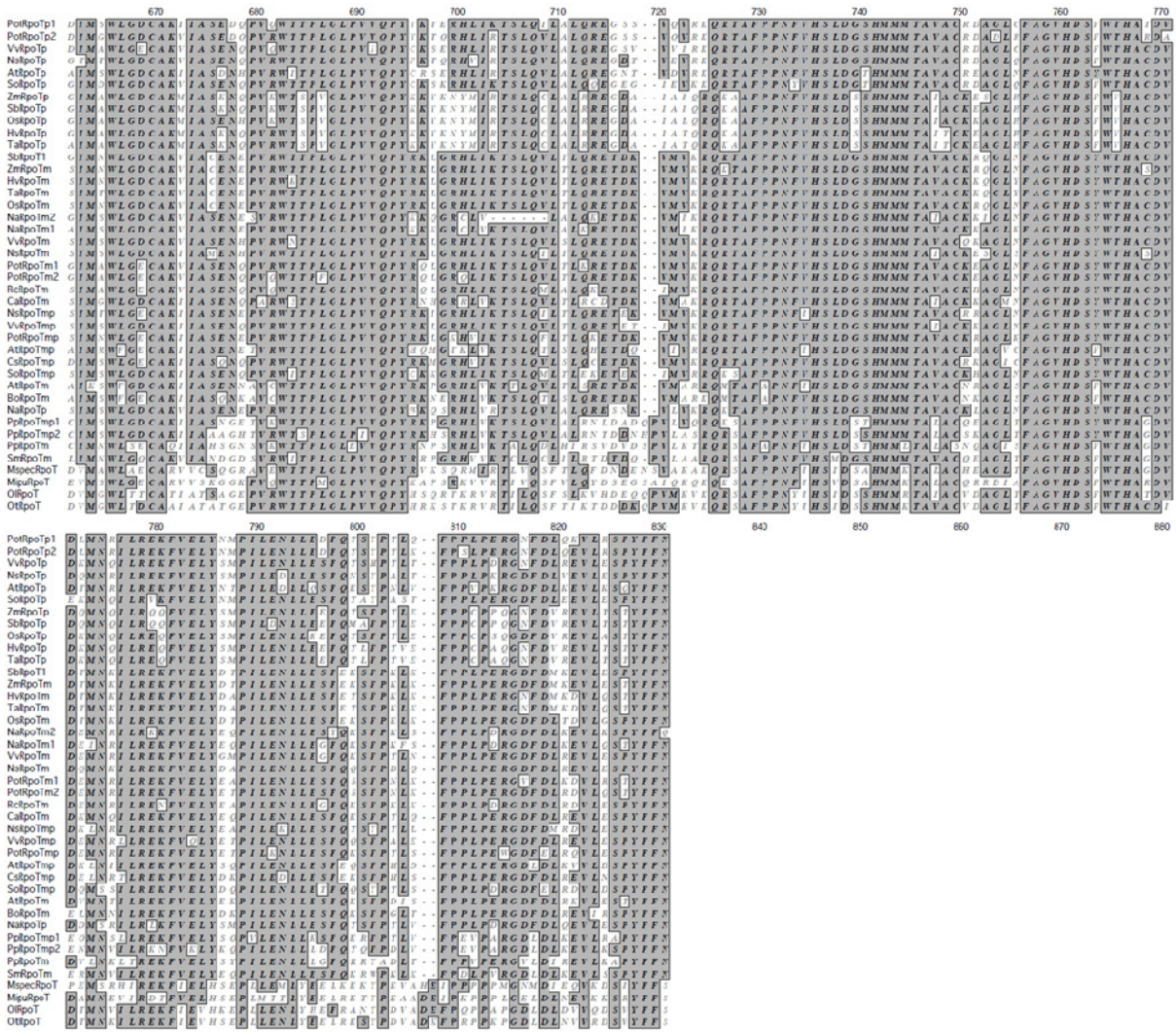
6.5 The partial cDNA sequence of *RpoT* gene from *Adiantum capillus-veneris*



6.6 Alignment of 41 phage-type RNA polymerases sequences







6.7 Abbreviations

aa	amino acid
A,C,G,T,U	nucleid acid bases (adenine, cytosine, guanine, thymin, uracil)
(d)ATP	(deoxy)adenosine triphosphate
BAC	bacterial artificial chromosome
BLAST	basic local alignment search tool
bp	base pairs
BSA	bovine serum albumin
°C	degree Celsius
cDNA	complementary DNA
CDS	coding sequence
CTAB	cetyl trimethyl ammonium bromide
(d)CTP	(deoxy)cytidine triphosphate
DNA	deoxyribonucleic acid
dNTPs	deoxy-Nucleotide TriPhosphates
DTT	dithiothreitol
EB	ethidium bromide
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tag
Fig	figure
g	gravitational acceleration
GFP	green fluorescent protein
(d)GTP	(deoxy)guanosine triphosphate
h	hour
IPTG	isopropyl-β-D-1-thiogalactopyranoside
kbp	kilobase pairs
kDa	kilodalton
min	minute
ML	maximum-likelihood
mm	millimeter
mM	millimoles per liter

MOPS	morpholinopropan-sulfonic acid
MP	maximum parsimony
mRNA	messenger RNA
µg	microgram
µl	microliter
µm	micrometer
µM	micromoles per liter
Na	*Nuphar advena*
NCBI	National Center of Biotechnology Information
NEP	nuclear-encoded RNA polymerase
nt	nucleotide
(d)NTP	(deoxy)nucleoside triphosphate
OD	optical density
PCR	polymerase chain reaction
PEP	plastid-encoded plastid RNA polymerase
PEG	polyethylene glycol
pH	potentia hydrogenii
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
RNAP	RNA polymerase
RpoT	RNA polymerases of the T3/T7 phage single subunit type
RT	reverse transcriptase
SDS	sodium dodecyl sulfate
sGFP	synthetic GFP
Sm	*Selaginella moellendorffii*
Tab	table
TAE	Tris-acetate EDTA
Tm	melting temperature
tRNA	transfer RNA
U	unit
UTR	untranslated region
UV	ultra violet

v/v	volume per total volume
w/v	weight per volume (g/100ml)
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

6.8 Acknowledgements

This dissertation represents the efforts of about five years of work, and would not have been possible to be finished without the effective cooperation and great assistance of so many people who deserve my acknowledgements.

I would like to express my highest gratitude to Professor Dr. Thomas Börner for his providing me the opportunity to conduct my PhD in the Department of Genetics, Humboldt University Berlin, for the guidance, for reviewing this thesis, for already being ready and kind to help me whenever I go to him. I learned a lot in his department, which paved the way for my future career development.

I am exceptionally grateful to our project leader, Dr. Andreas Weihe, for his in-detail supervision, strong encouragement and full supports for my PhD study, for his careful correction of this dissertation, and for his translation of the summary from English into German as well as for his advice on my career development which is of importance for my future.

A special thank-you goes to Mr. Uwe Richter for his stimulating discussions and sharing his ideas with me, especially for his work on isolation of some *Na* BAC DNAs, the alignment of 41 RpoTs and the construction of Mr. Bayes tree. Susanne Beick and Björn Richter also contributed in this work.

Thanks to Prof. Dr. Christian Schmitz-Linneweber and Dr. Michi Tillich who helped me for the work of protoplast isolation and PEG transformation. Dr. Christina Kühn deserves my thanks for introducing me the con-focal microscopy. Sequencing was carried out by Dr. Martin Meixner (SMB GmbH, Berlin). Thanks are also due to all my co-workers (Dr. Karsten Liere, Cornelia Stock, Petra Dreier and Eva Hackenberg) for their efficient daily lab support, for their warmth on my numerous requests for German translation during my staying in the lab, and for a lot of very generous help.

My friend, Dr. Xiaohua Chen provided me generous help with both the living of my family in Berlin and my PhD study in Humboldt University Berlin, which I would like to cherish in my heart.

Finally, I want to express my deep appreciation to my families: my mother has been providing me with her continuous support over all these years which enable me to get this far; my husband Dr. Jianping Liu, with his love and care, has made my life in the lab full of joy and interests as at home; my son Yuchong Liu and my daughter Yuchen Liu, with their cute smiles as well as the bothering cry during the years of growing up, clears up and fuels my mind every time when I stay with them.

My PhD study was generously and partially supported by NaFöG. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (WE 1595/6-2) to Professor Dr. Thomas Börner and Dr. Andreas Weihe.

6.9 Publications

6.9.1 Publications in peer-reviewed international journals

1. **Chang Yin**, Uwe Richter, Thomas Börner, Andreas Weihe. Evolution of plant phage-type RNA polymerases: The genome of the basal angiosperm *Nuphar advena* encodes two mitochondrial and one plastid phage-type RNA polymerases. (BMC Evol Biol. 10:379. 2010 Dec 6.).
2. **Chang Yin**, Uwe Richter, Thomas Börner, Andreas Weihe. 2009. Evolution of phage-type RNA polymerases in higher plants: Characterization of the single phage-type RNA polymerase gene from *Selaginella moellendorfii*. Journal of Molecular Evolution. 68(5):528-38. Epub 2009 May 1.

6.9.2 Publications in Chinese

3. Jianping Liu, **Chang Yin.** Book translation from English into Chinese, Introduction to health services (Edition Six) (Authored by Stephen J. Williams and Paul R. Torrens), 2004, published by Medical Publishing of Peking University and Peking University publishing
4. Jianping Liu, Aiguo Xin, **Chang Yin.** The Distribution of NOS Positive Nerve in Mice and the Biological Functions of Nitrogen Monoxide, 2001, 31(2), Gansu Animal and Veterinary Sciences (Review in Chinese)
5. **Chang Yin**, Jinghua Fan, Advance and Prospect of Biosensor Research on Environmental Analysis, 2000, 6, Biotechnology Information (Review in Chinese)
6. Jianping Liu, **Chang Yin**, Gefen Yin, Advances on Prion, 2000, 21(3), Progress in Veterinary Medicine (Review in Chinese).

6.10 Erklärung

Ich versichere, dass ich die vorliegende Arbeit im Fach Biologie selbständig, ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe.

Die dem Verfahren zugrunde liegende Promotionsordnung ist mir bekannt.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Berlin, den 27.07.2010